(continued from part 32)

Three basic systems

In conventional broadcast television there are presently three basic systems of creating, transmitting, and receiving colour images: the American **NTSC** system; the German **PAL** system (derived from the NTSC system and now used in most of Europe and the U.K.); and the French **SECAM** system. We'll look at all three systems in closer detail later, but for now we'll restrict ourselves to the U.K. PAL system, which is a 625 line standard.

The PAL system comprises a scan speed of 15,625 lines each second, creating 50 fields a second; that is, a 15,625 Hz horizontal scan frequency and a 50 Hz vertical scan frequency. The electron beam therefore scans each line and returns to the next in 64 μs. In this period, the visible line on the screen itself is actually scanned in 52 μs; the other 12 μs is taken up by the **line flyback**, during which the electron beam returns from the end of one line to the beginning of the next.

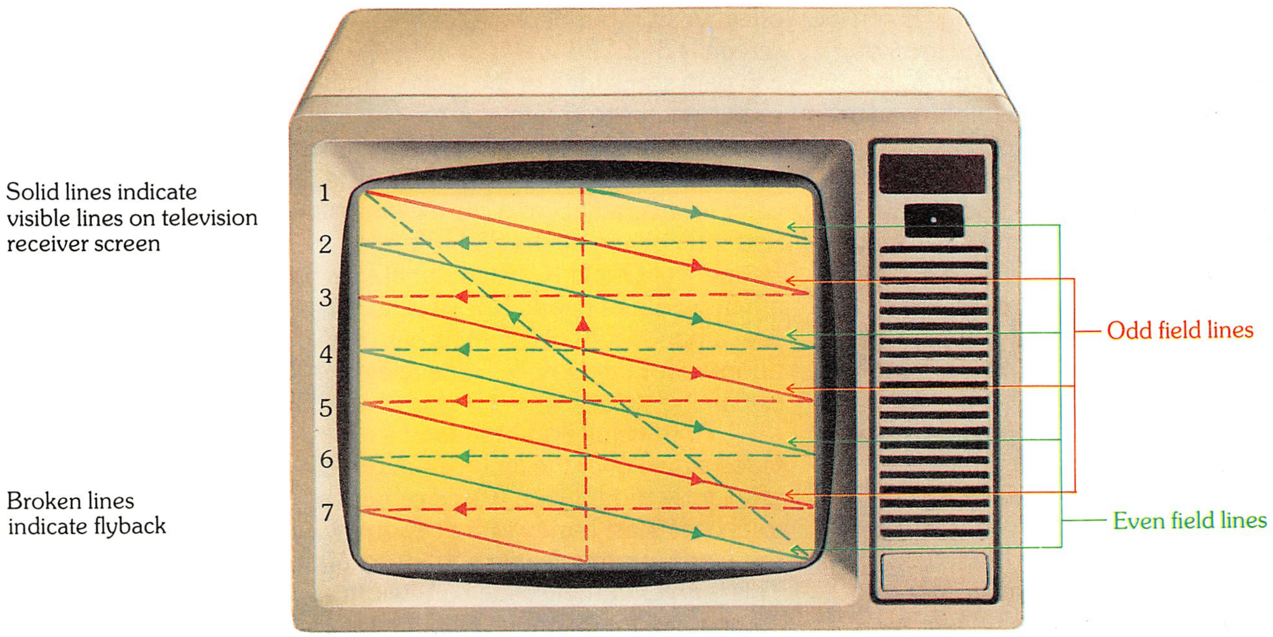
If 15,625 lines are scanned each second (and the U.K. PAL system operates at 50 fields a second), it appears that only 312.5 lines are scanned for each field. We have just said that the U.K. was a 625 line standard, so why are only 312.5 lines scanned in a field?

Interlaced scanning

The answer to this question is that it takes two complete scans from top to bottom of the camera screen to form one apparent 625 line image on the television receiver screen. Each complete television picture is composed of *two* vertical scans – each having half the number of lines of the complete picture. A complete television

5. A television screen illustrating the principle of interlaced scanning.

5



picture is produced once every 1/25 second and these two fields are said to provide one **frame** of an interlaced picture.

The first field starts at the top left-hand corner of the screen and scans all the odd numbered lines finishing at the bottom centre of the screen. The beam flies back to start the second field from the centre top of the screen, interlacing the even lines with the impression of the last field's odd numbered lines. The second field finishes at the bottom right-hand corner of the screen and the odd numbered line scan begins again. This dual process takes 1/25 of a second and scans all 625 lines. A seven-line, two-field, interlaced scan television screen illustrating this principle is shown in *figure 5*.

You may ask why two interlaced fields are necessary to develop one picture? Surely a full 625 line picture could be produced in a single field.

The reason is that the transmission bandwidth must be kept as narrow as possible. The horizontal definition of the picture is influenced by the speed of the scanning beam. The faster the beam, the less time there is for the brightness level to change and show the picture detail on the screen. But, the brightness change is controlled by the speed at which the level of the vision signal can change in the television system as a whole, and this is a factor of system bandwidth. Any increase in horizontal scanning speed demands an increase in bandwidth. The additional bandwidth required by a single-field-per-frame television system would be acceptable if there was plenty of free air space. But there isn't! The use of a two-field-per-frame television system enables a more efficient utilisation of available air space.

Perhaps to produce a one-field-per-frame television system, with the same bandwidth as the present two-field-per-frame system, the vertical scan frequency (the **field frequency**) could be slowed to 25 Hz and the 15,625 Hz horizontal scan frequency (the **line frequency**) retained? This would eliminate the extra bandwidth requirement, but it would produce picture flicker. The field frequency has to be at least 50 Hz before the flicker is acceptable.

Interlacing solves both these problems and allows the transmission and reception of acceptable quality pictures, within a practical bandwidth, at a flicker-free field frequency.

Pixels

A **pixel** (or **picture element** or **pel**) is the term applied to the smallest addressable point in a graphics display – which could be a television screen or a VDU. Suppose we have a 625 line picture on a screen whose height equals its width. Not all the lines are visible: in the U.K. system, for example, about 20 lines in each field are inactive or used for purposes other than the transmission of a television picture. So, with 585 **active** lines in this square picture, there are 585 × 585 pixels, i.e. 342,225. The horizontal resolution of the resulting picture is equal to the vertical resolution.

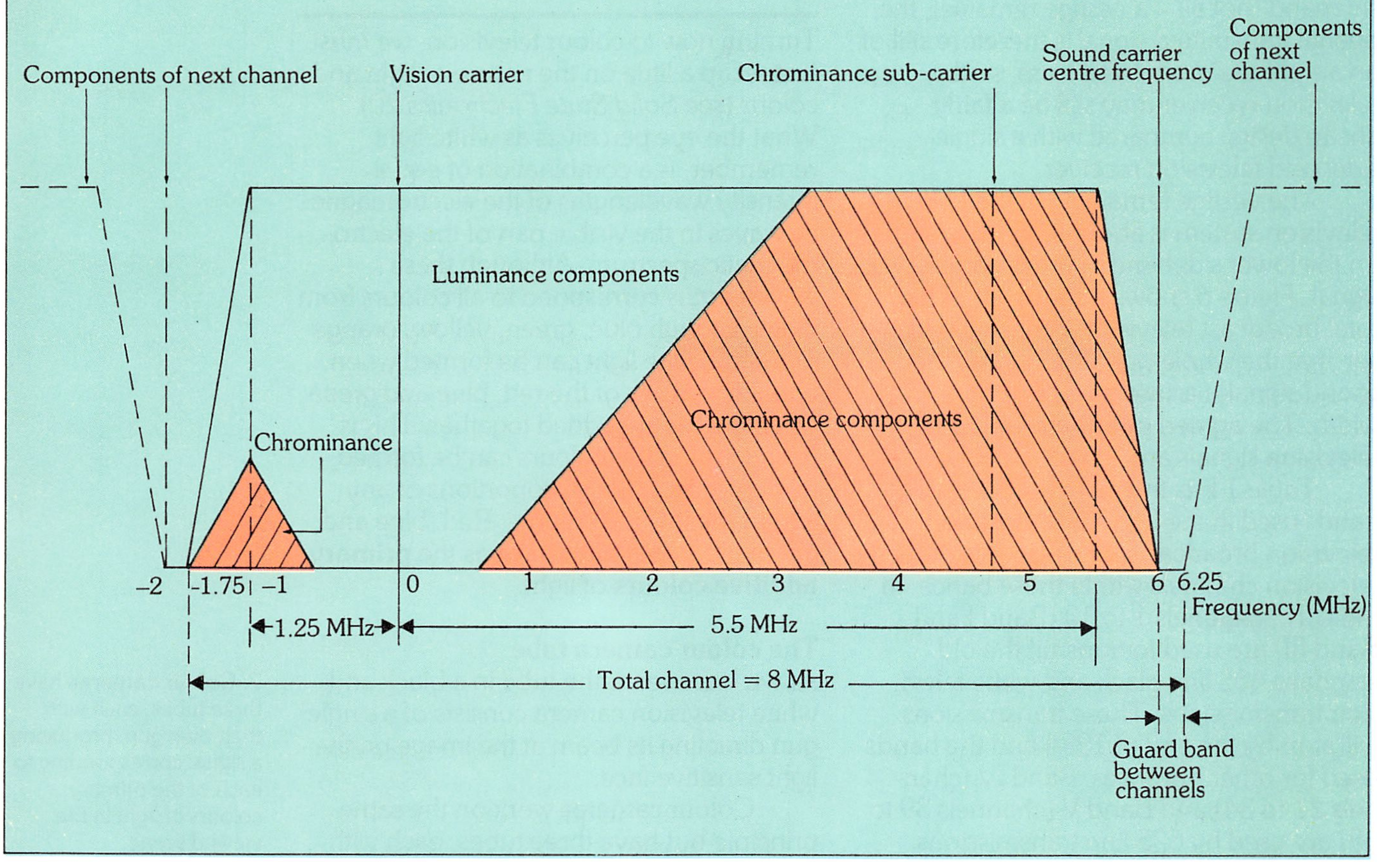
This, of course, assumes the scanning beam produces a displayed spot, whose diameter is equal to the width of a scanning line, i.e. the height of the screen divided by 585. The spot size must, in fact, be such that it does not overlap any of the lines or leave gaps between them.

The **aspect ratio** of current television sets is not 1:1, but 4:3, i.e. the width is four units while the height is three units. The effective number of pixels for *this* shape of screen is 342,225 × 4/3 or 456,300. We can gain some idea of the video bandwidth requirements to create a picture of this number of pixels by halving the number (a general rule of thumb) – as each cycle may cater for *two* adjacent picture elements – and multiplying by the 25 Hz frame frequency. Our U.K. video signal thus has a bandwidth of nearly 5.7 MHz. In practice, a bandwidth of 5.5 MHz has been adopted for the U.K. 625 line television signal.

To transmit and broadcast such a video signal over the air by envelope amplitude modulation would require a total broadcast bandwidth of about twice this (see *Communications 3*), or 11 MHz. This is an extremely wide bandwidth – compare it, for example, with the total available frequency on the medium wave radio band of only 1 MHz – and we can therefore get some idea of the problems involved in broadcasting television signals.

Technically, as with telephone communications (see *Communications 3*)

6



6. A spectrum of the total broadcast television signal showing that it fits into an 8 MHz bandwidth.

Table 1
Main radio and television broadcasting bandwidths in the U.K.

Band	Wavelength	Channel	Use
Low frequency	160 to 225 kHz	Long wave	AM radio
Medium frequency	525 to 1605 kHz	Medium wave	AM radio
Very high frequency			
Band I	41 to 68 MHz	1 to 5	405 line television
Band II	88 to 97.6 MHz		FM radio (stereo)
Band III	174 to 216 MHz	6 to 13	405 line television
Ultra-high frequency			
Band IV	470 to 582 MHz	21 to 34	625 line colour TV
Band V	614 to 854 MHz	39 to 68	625 line colour TV

single sideband modulation could be used to transmit the signal, with an overall broadcast bandwidth equal to the signal bandwidth. This is possible and cost effective for wired communications and speech communications, but it becomes more difficult for broadcast communications and even more difficult for video communications. It is also much less cost effective: the cost of each television receiver would be greatly increased if it had to receive single sideband modulation transmissions.

In order to keep receiver costs as low as possible and to reduce the required broadcast bandwidth of the transmitted signal, a different type of transmission is used, known as **vestigial sideband modulation**. This operates on the principle that

each sideband of an envelope amplitude modulated signal contains identical information. By filtering out *some* of one sideband (not all – a *vestige* remains), the overall transmitted signal is therefore still of an amplitude modulated form, so the television receiver may still be a fairly cheap design compared with a single sideband television receiver.

The vestige remaining in the U.K. television system is about 1.25 MHz, and is on the lower sideband of the broadcast signal. *Figure 6* shows a spectrum of the total broadcast television signal and we can see that the whole signal (including the sound signal) fits into an 8 MHz bandwidth. The agreed **channels** for U.K. television signals are, in fact, 8 MHz wide.

Table 1 lists the main broadcasting bands used in the U.K. for radio and television broadcasts, and also lists the television channels within these bands. In the U.K., channels 1 to 13 (Band I and Band III) are used to transmit the old standard 405 line black and white television transmissions. These transmissions will stop by the end of 1984 and the bands used for other purposes. Band IV (channels 21 to 34) and Band V (channels 39 to 68) are used by 625 line transmissions.

Colour television cameras and receivers

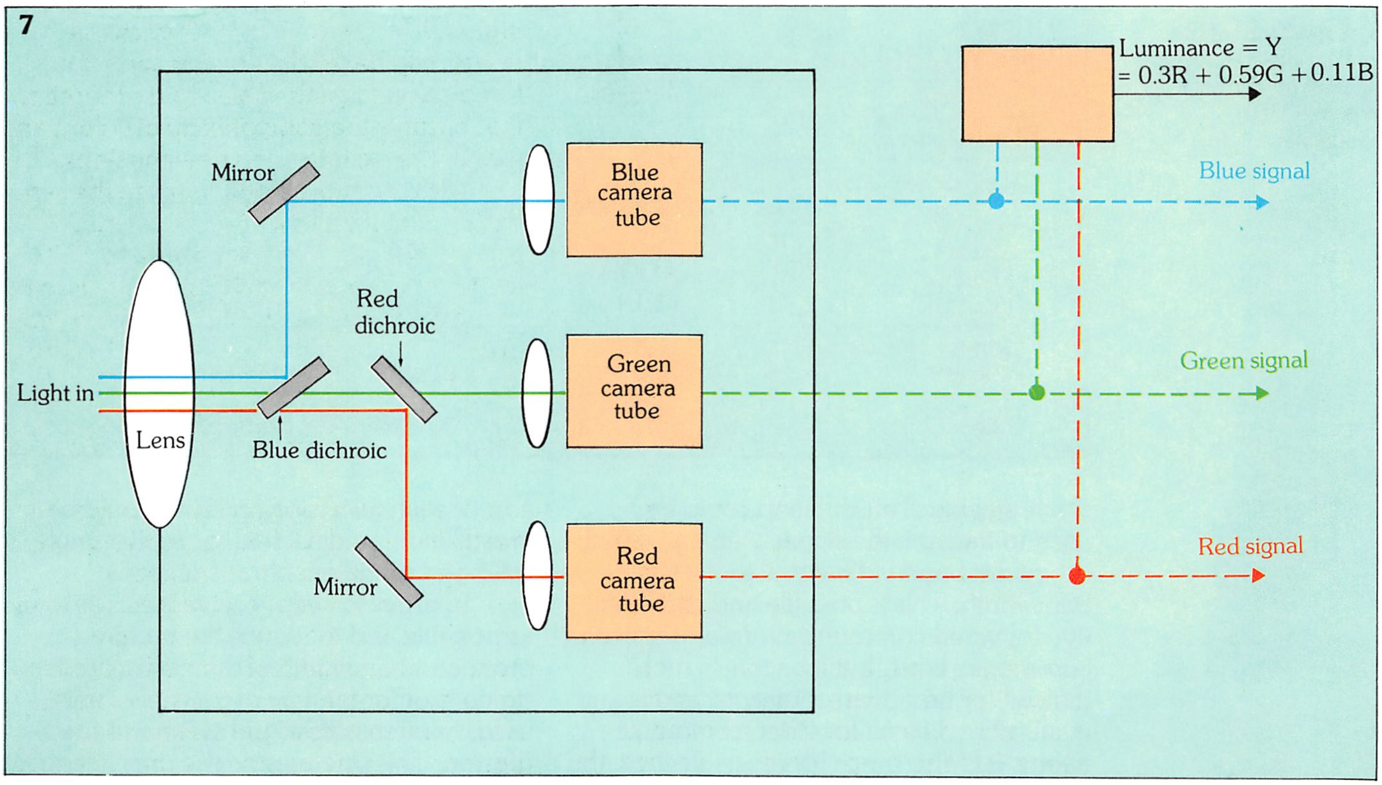
Turning now to colour television, we must first recap a little on the nature of light and colour (see *Solid State Electronics 20*). What the eye perceives as white light, remember, is a combination of equal intensity wavelengths of the electromagnetic waves in the visible part of the electromagnetic spectrum. Although these wavelengths correspond to all colours from violet through blue, green, yellow, orange and red, white light can be formed when equal quantities of the red, blue and green wavelengths are added together. This is because all other colours can be formed from mixing various proportions of any two of these three colours. Red, blue and green are therefore known as the **primary additive colours** of light.

The colour camera tube

As we have seen, the tube in a black and white television camera consists of a single gun directing its beam at the image on the light sensitive face.

Colour cameras work on the same principle but have three tubes, each with

7. Colour cameras have three tubes, each with their own gun, producing a signal corresponding to each of the primary colours of light in the viewed scene.



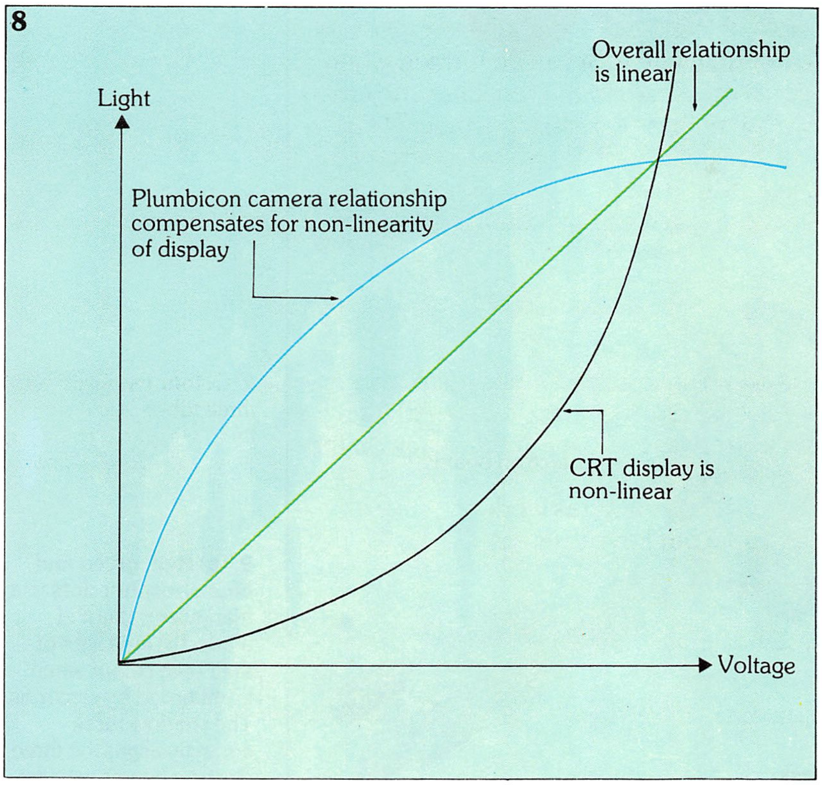
their own gun, producing a signal corresponding to each of the primary colours of light in the viewed scene. A block diagram of such a camera is shown in *figure* 7.

Special coloured mirrors known as **dichroics** split the light passing from the scene being televised into its red, green and blue constituents. After passing through the camera lens, the light first strikes a blue dichroic mirror which reflects only the blue component of the image. This blue component then strikes an ordinary mirror and is focused through another lens onto the blue camera tube. The output from this tube than becomes the blue signal.

The red and green light passes through the blue dichroic mirror; the red light is then reflected by the red dichroic mirror onto the red camera tube, thereby becoming the red signal, and the green light, which passes through the red dichroic mirror, is focused onto the green camera tube and becomes the green signal.

8. A plumbicon camera produces an overall linear relationship between the light levels produced on the television receiver and the light levels sensed by the television camera.

8



Gamma correction

It may be thought that the light levels produced on the television receiver screen should be linearly proportional to the light levels sensed by the television camera. However, this is not the case because a cathode ray tube (in television receiver *and* camera) is a non-linear device.

To compensate for these non-linear relationships the signals from the camera are **gamma corrected** by using a special type of camera – a **plumbicon** camera. *Figure 8* shows how the voltage-to-light ratio of the plumbicon camera corrects for the voltage-to-light ratio of a television display tube, to produce an overall linear relationship.

Luminance and chrominance

Colour has three independent characteristics: brightness, more correctly called **luminance**; **hue**, the colour's dominant wavelength; and its **saturation** or strength.

The three colour signals from the three tubes of the colour camera may be mixed in certain proportions, to provide a monochrome (black and white) signal which, after transmission and reception, may be used by a monochrome television receiver to create a monochrome picture. Monochrome television pictures, as we have seen, are created by varying the light intensity produced by the spot of a single electron beam – we can therefore see that this monochrome signal is a **luminance signal**. It is the fact that the luminance signal *may* be produced by mixing the three colour signals in a certain proportion, that ensures the compatibility of colour and monochrome television transmissions.

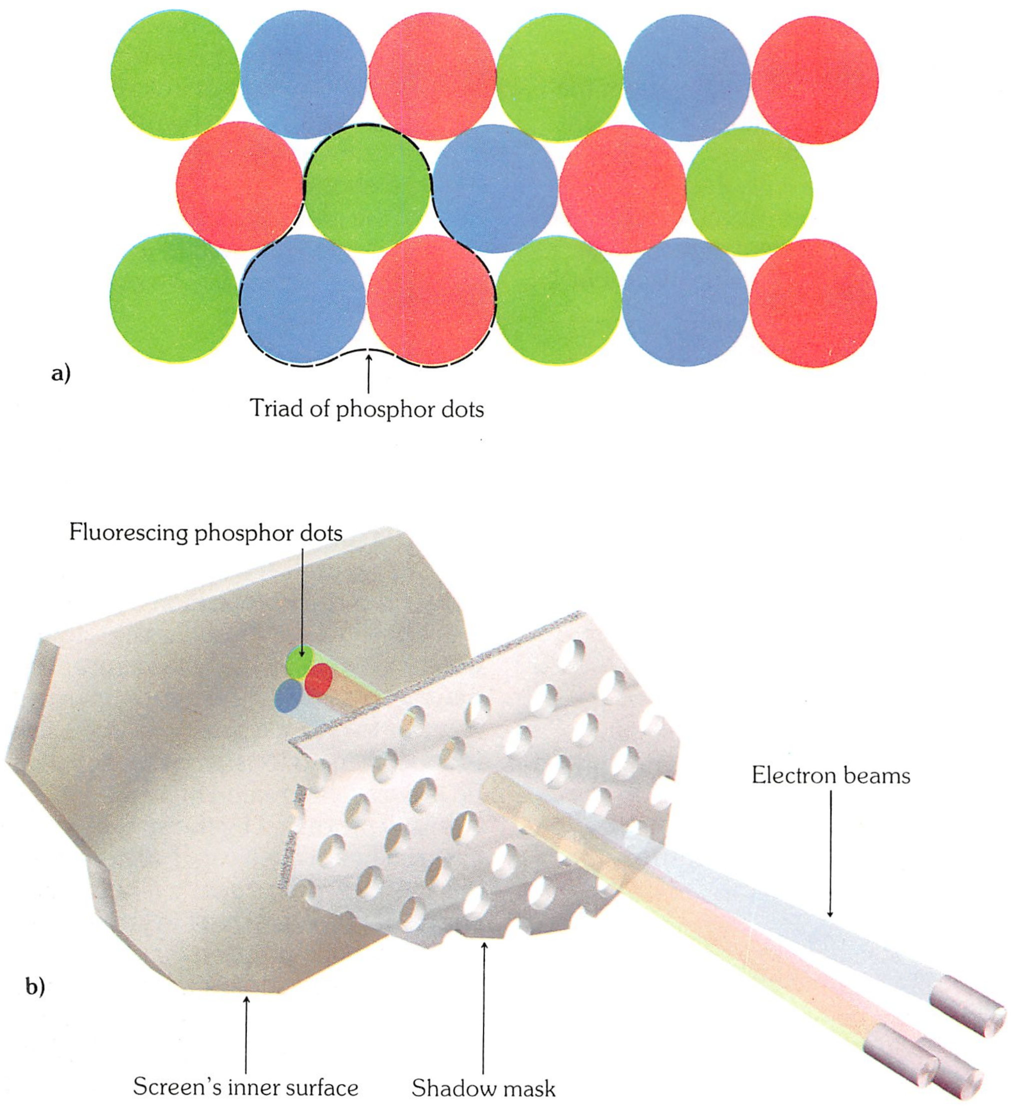
The luminance signal (given the symbol Y) is built up from 30% of the red signal, 59% of the green signal and 11% of the blue signal. When all three signals are of equal magnitude, this combination produces white, and may be summarised by the formula:

$$Y = 0.3R + 0.59G + 0.11B$$

where R, G and B represent the three colour signals.

To enable colour television receivers to recreate a colour picture, *separate* colour signals are modulated onto this luminance signal. Since the luminance signal is an algebraic function of three colour signals, an additional two more signals need to be sent for the receiver to work out the relative intensities of the three colour beams and thereby produce the

9



9. (a) Red, green and blue phosphor dots in a triangular pattern of triads; **(b)** emit light of their own colour when bombarded by electrons. The shadow mask correctly aligns the three electron beams with the corresponding phosphor dots of each triad.

colour image.

The two most suitable signals to send are known as the **colour-difference** or **chrominance signals**: R-Y, the colour difference between the red signal and the luminance signal; and B-Y, the colour difference between the blue signal and the luminance signal. It is the red and the blue colour-difference signals which are transmitted because the green-difference (G-Y) is the smallest of the three (the green *signal* is largest) and would therefore be more susceptible to transmission problems and interference.

In the television receiver, a simple resistor network derives all three colour levels from the two chrominance signals and the luminance signal.

The luminance signal is the signal which is transmitted as a vertigial sideband modulated signal. The chrominance signals are envelope amplitude modulated onto a **sub-carrier** signal, which has a frequency of 4.43361875 MHz above the main carrier frequency. In fact, two sub-carriers are used, one for each colour difference signal, with the same frequency but 90° apart in phase. They are added together to produce a **single phase** or **quadrature modulated** sub-carrier. Before

transmission the sub-carrier frequency signal is suppressed and the chrominance information carried purely on its sidebands. The signal that results is the vector sum of the chrominance signals and represents: by its phase, the dominant hue of the original light; and by its amplitude, the colour's saturation. It has a broadcast bandwidth of about 2 MHz.

Television receivers

We have previously seen in *Solid State Electronics 23* that the cathode ray tube in a television receiver is rather like a large glass envelope, the inside face of which is coated with phosphor which glows when bombarded with electrons. As they stream towards the screen, these electrons build up kinetic energy (just like the head of a hammer on the swing towards a nail) which is then liberated on impact with the screen and transformed into light radiation.

In monochrome CRTs, white light is omitted from monochromatic phosphor. In a colour CRT, the picture is built up from a specific pattern of phosphor dots. These separate red, green and blue phosphor dots, arranged in a triangular pattern of **triads**, as shown in *figure 9*, emit light of their own colour when bombarded by electrons.

The electrons are fired from three separate electron guns which focus their respective beams onto the tube so that only their own colour phosphor is hit.

A **shadow mask** – a plate perforated with a hole for each triad of colour dots – is placed between the electron beams and phosphor to direct each beam specifically to its own colour phosphor dot. Even when the beam is deflected it will still hit the same colour dots. It obviously takes a great deal of accuracy to produce colour television tubes. On a television screen, for example, there could be over 400,000 holes in the shadow mask and therefore at least 1.2 million phosphor dots.

If a beam does not fall on a phosphor dot of the same colour, it hits the shadow mask which absorbs up to 80% of the beam's energy. In terms of light output for a given beam current, we can understand that colour CRT is less efficient than the monochrome tube. (See *Solid State Electronics 23*.)

Keeping the picture in step

To perfectly reproduce the televised scene, each line displayed on the receiver screen must be recreated in exactly the right position, so the electron beams in the receiver must move in step with the camera.

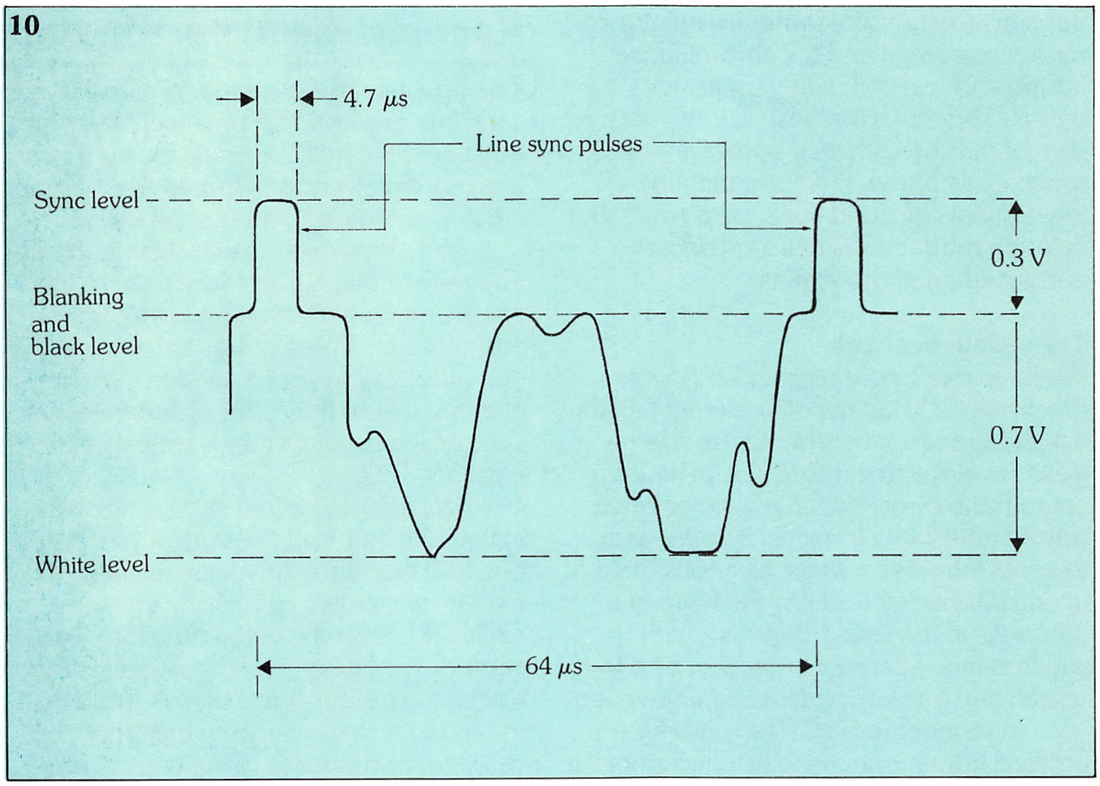
We already know that signals known as **synchronising pulses** are used to initiate the beam deflections in both camera and receiver. These pulses are, in fact, generated in the studio and fed into the camera, where they control timebase circuits which produce the beam deflection currents.

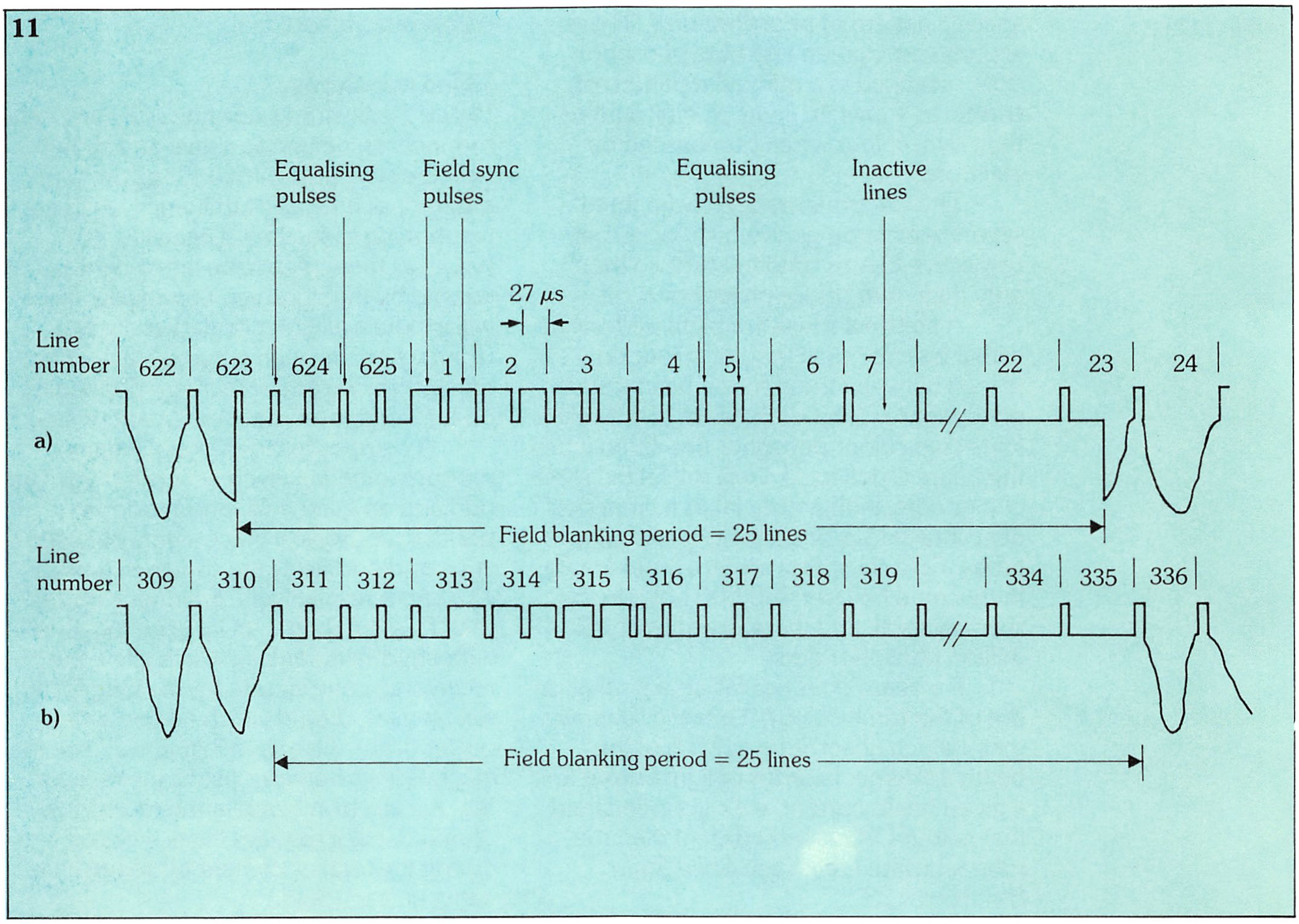
We also know that short sync pulses initiate the timing of the horizontal deflection and therefore the timing of each line; longer sync pulses control the timing of each field. After the video signal leaves the camera, the line and field sync pulses (and a separate **colour burst** signal – used by the receiver to synchronise the reference oscillator and to identify the two colour-difference signals) are added, forming the composite video signal.

Video waveforms

Figure 10 illustrates one line of a composite monochromatic video signal and shows two line sync pulses (of 4.7 μs duration) and monochromatic luminance levels corresponding to the beam intensity as it traverses the screen from left to right as viewed by the observer. Obviously, this will produce a lighter or darker display (from white, through grey, to black on a monochrome receiver screen) depending on the voltage level at any instant of time.

A composite video signal (again monochromatic) which occurs between an odd and an even field, and the corresponding composite video signal between even and odd fields are shown in *figures 11a* and *b* respectively. Field sync pulses (of 27 μs duration) can be seen together with **equalising pulses** which allow the receiver to compensate for the long field sync pulses, during which no line sync pulses occur. Also seen in *figure 11* are the inactive lines between fields: these show that the electron beam in the receiver is **blanked**, i.e. it displays black – it is off during the total of 25 **field blanking** lines between each field.

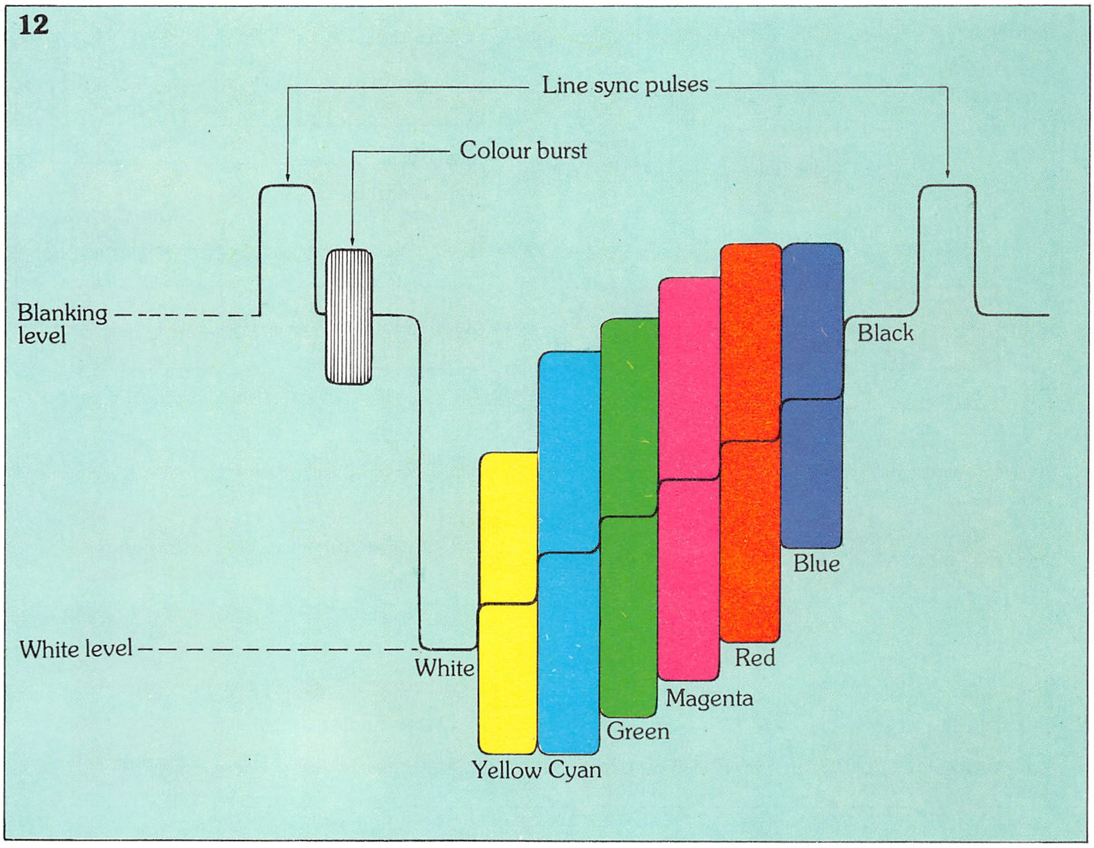




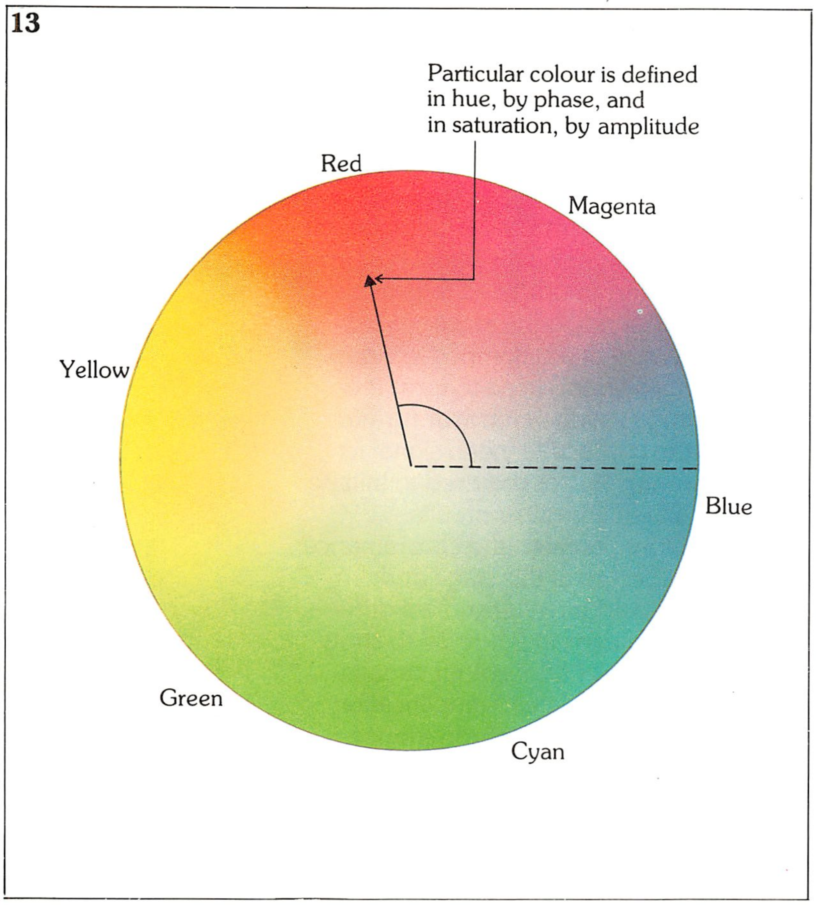
10. One line of a composite monochromatic video signal.

11. (a) A composite monochromatic video signal occurring between an odd and an even field; **(b)** the composite video signal between an even and an odd field.

12. The composite video signal of a single line of the standard vertical colour bar chart.



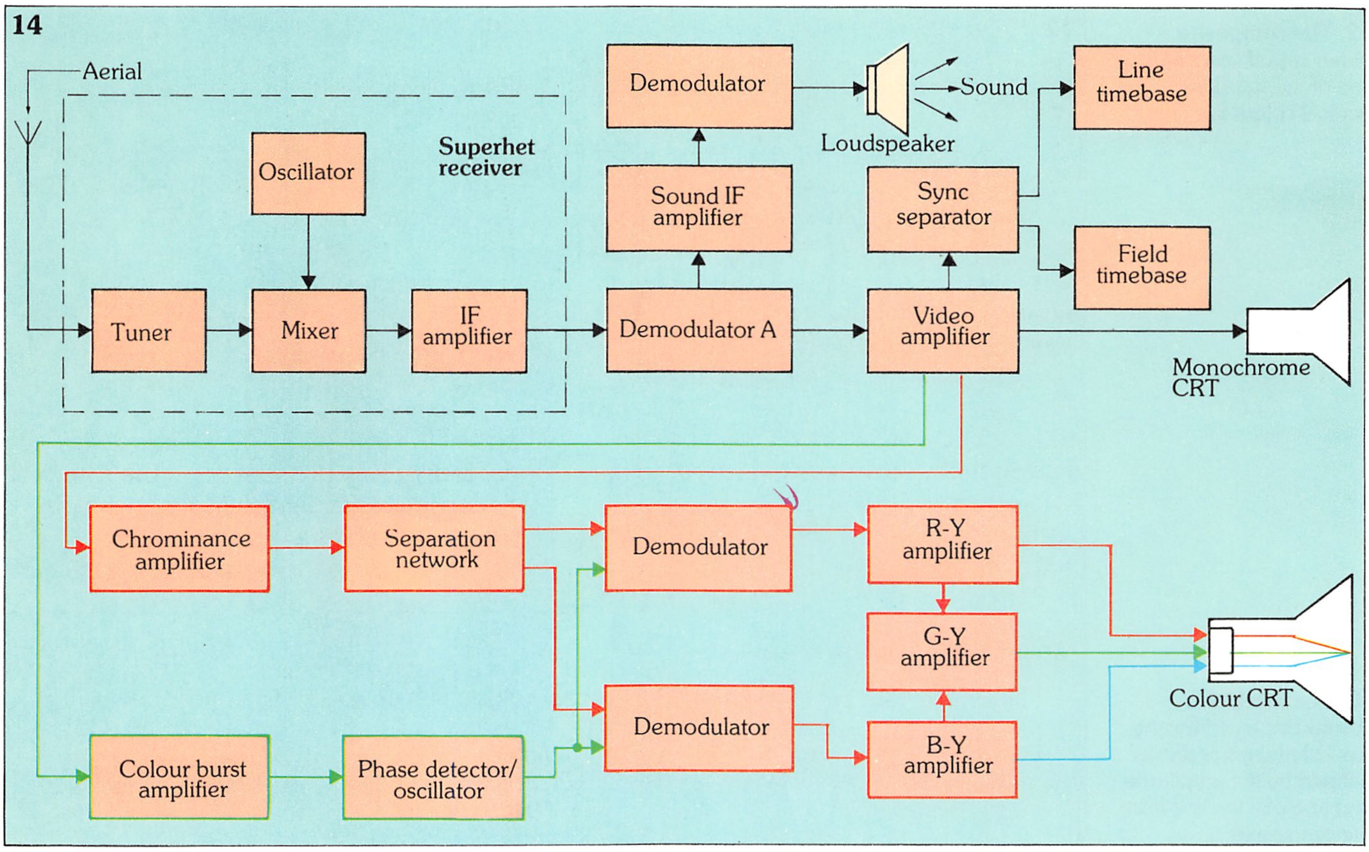
13. The particular colour displayed by any triad of phosphor dots is defined by the amplitude and angle of the line from the centre axis.



The composite video signal of a single line of the standard vertical colour bar chart is shown in *figure 12*. The colour burst signal is shown, as are the reference levels, to enable a colour receiver to display the eight colours (white, yellow, cyan, green, magenta, red, blue and black). This composite video signal contains all the information required to move the three electron beams of a colour CRT across the screen and to modulate their relative intensities, producing the colours.

The composite video signal waveforms shown in *figures 10, 11* and *12* illustrate how a greater voltage produces an electron beam of lesser intensity and therefore a darker display. Similarly, a lower voltage produces a brighter display. For this reason we say that the video signal is **negatively modulated**.

Positive modulation, on the other hand, would occur if the beam becomes more intense as the voltage increases. Some television systems use positively modulated signals (the old 405 line U.K. system, for example) but there is a significant advantage in the use of negative modulation, in that electrical interference is



14. Block diagram of a PAL television receiver.

not as noticeable.

The amplitude and phase of the chrominance signals determine the currents for the three beams of the colour CRT. Amplitude contains the information relating to colour saturation, while phase contains the information relating to colour hue. This may be represented in a colour circle as shown in *figure 13*. Here, the particular colour displayed by any triad of phosphor dots on the screen is totally defined by the amplitude and angle of the line from the centre axis.

PAL colour receiver

A block diagram of a PAL television receiver is shown in *figure 14*. The coloured parts of the diagram along with the three-gun CRT form the only differences between monochrome and colour receivers. We shall therefore begin with the operation of a monochrome receiver.

The **tuner**, **oscillator**, **mixer** and the **intermediate frequency** (IF) **amplifier** receive the broadcast television signals and, regardless of the exact transmission frequencies, turn them into two *common* intermediate frequencies. In the U.K., these are 39.5 MHz for vision and 33.5 MHz for sound, i.e. they are 6 MHz apart. A combination of circuits such as this is known as a **superhet radio receiver**.

Returning to *figure 14*, demodulator A separates the sound signal from the video signal but leaves it still modulated on a 6 MHz carrier signal. This signal is amplified and further demodulated to provide a sound output signal to the loudspeaker.

The demodulated composite video signal is amplified and used to form the beam current. A **sync separator** further processes the composite video signal to provide the sync pulses required to initiate the line and field timebase circuits.

In a colour receiver, the chrominance signals are detected at this stage and passed to the colour circuitry. Two different colours are used in *figure 14* to show the main functions of the colour circuits. Those parts in red relate to amplification and demodulation of the chrominance signal into the colour signals for the three guns; those parts in green relate to the regeneration of the sub-carrier frequency necessary to demodulate the chrominance signal.

Transmitting the signal

The public broadcast television network of UHF and VHF 625 line transmissions, using the PAL system, now covers some 99% of the U.K.; transmitters provide coverage for isolated villages.

Transmitter networks in the U.K. are planned on the basis that a guaranteed minimum field strength will be obtained using receiving aerials with adequate sensitivity and directivity. With an 8 to 10 element UHF receiving aerial, and average cable losses, for example, this minimum field strength provides signal input to the receiver of about 750 μV in Band IV, decreasing to about 350 μV in the higher frequency part of Bank V.

Other colour television systems

The first practical colour television broadcasting system was developed in the United States in 1953, after a variety of methods had been proposed and tried. It was devised to be fully compatible with the American 525 line, 60 Hz monochrome system. Known as the NTSC system (after the National Television System Committee), most of its ideas were adopted by the other two major systems, PAL and SECAM.

The PAL system adopted for the U.K. was originally proposed in Germany and later adopted throughout Europe. PAL (Phase Alternating Line) describes the way in which the colour system is encoded into the television signal, one colour-difference signal per alternate line.

The SECAM system (SÉquential Couleur À Mémoire) was adopted by France, the USSR and some of the East European countries.

The major differences between PAL, NTSC and SECAM

The major difference between the three systems is the way in which they combine the two chrominance signals and superimpose them onto the luminance.

Like PAL, the NTSC system also forms two chrominance signals and amplitude modulates them onto a sub-carrier (at 3.58 MHz) within the luminance signal.

However, the way these signals are added to the luminance signal creates problems. Any phase shifts in the received signal are immediately visible on the displayed picture as a change in hue.

The French system, SECAM, is known as a sequential system transmitting the colour-difference signals *sequentially on alternate lines*. The phasing of the received signals therefore does not matter. In the receiver, a delay system ensures that the signals are processed at the same time. SECAM enables receiver design to be more simple and reduces colour distortion due to phase problems. Studio work, however, is more complicated.

PAL is basically similar to NTSC but is designed to overcome phase shift problems by reversing the phase of the chrominance signal on alternate lines. In the receiver, any phase shifts are thus averaged out, and changes in hue drastically reduced, if not eliminated.

In the PAL receiver, a delay line in the decoder delays the incoming signal by 64 μs (one screen line) and then compares it with the original; the output is the average of the two signals (see *Table 2*).

Table 2
Main properties of the PAL, SECAM and NTSC systems

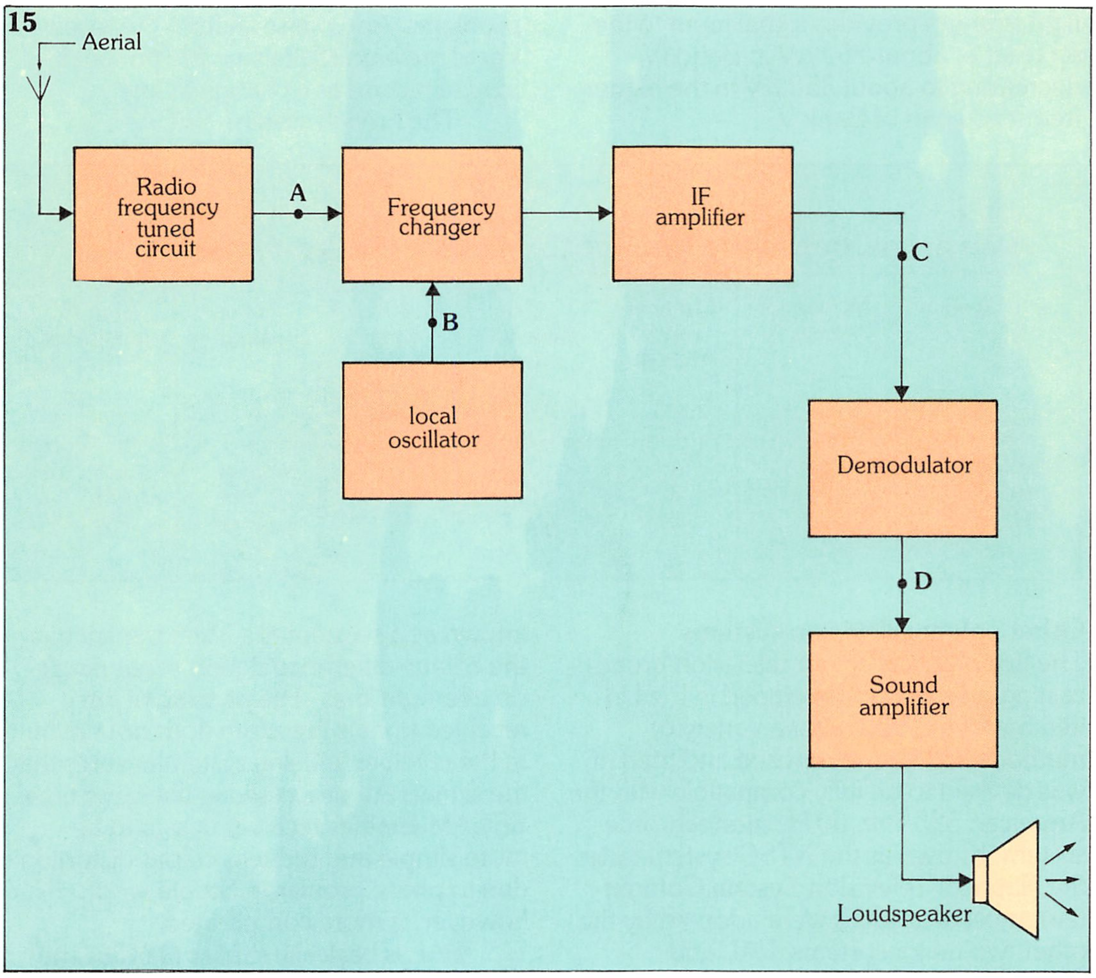
System	Number of lines	Channel width (MHz)	Video bandwidth (MHz)	Position of sound carrier from video carrier	Vision modulation	Sound modulation
SECAM	819	14	10	–11.15	negative	AM
PAL	625	8	5.5	+6	positive	FM
NTSC	525	6	4.2	+4.5	positive	FM

Radio receivers

Most modern radio receivers operate in a similar way to the sound circuits in television receivers. In a typical radio receiver, shown in block diagram form in *figure 15*, the radio frequency signal is selected by a radio frequency tuned circuit and applied, together with the output of an oscillator known as the **local oscillator**, to a circuit known as a **frequency changer** (or **mixer**).

The output of the frequency changer is a signal which is modulated with the termediate frequency, the frequency of the signal generated by the local oscillator must be variable, so that as different radio transmissions (of different carrier frequencies) are received, the *difference* between the carrier frequency and the local oscillator frequency remains the same.

An intermediate frequency amplifier with a very narrow bandwidth centre on the intermediate frequency amplifies the signal. Because the amplifier has such a narrow bandwidth, other radio transmissions with a carrier frequency close to the

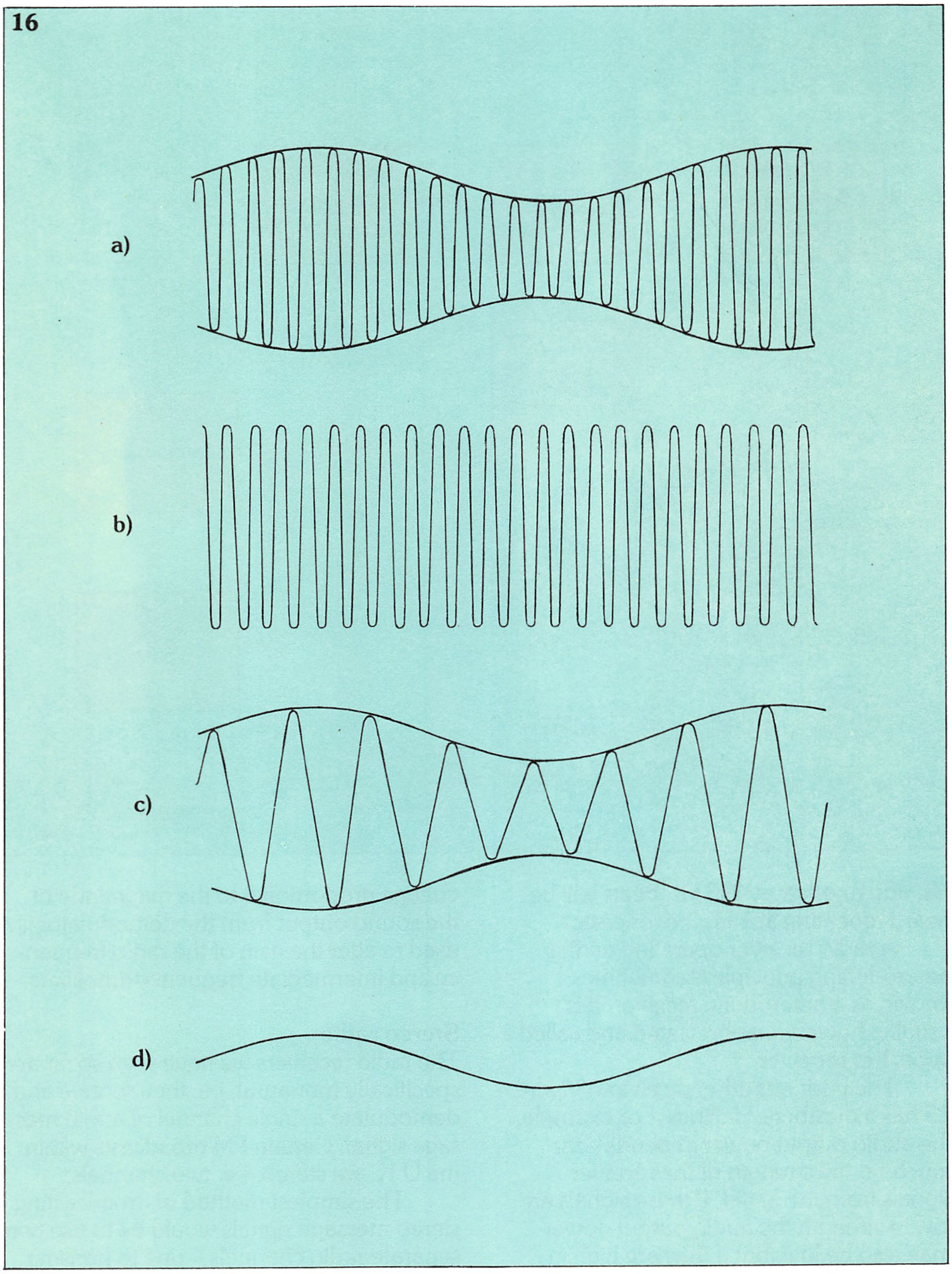


15. Block diagram of a typical radio receiver.

originally transmitted message signal, but *at a different frequency*. This different frequency is, in fact, the *difference* between the originally received carrier frequency and the oscillator frequency. It is known as the **intermediate frequency**, and is constant (generally about 465 to 468 kHz in AM receivers and 10.7 MHz in FM receivers).

In order to ensure a constant in- one received are *not* amplified. Such radio transmissions are often called **adjacent channels** and, if they *are* amplified so that they interfere with the intermediate frequency, they produce what is known as **adjacent channel interference**.

The amplified intermediate frequency signal is demodulated by a demodulator, often called a **detector**, which produces the audio message signal output. This is

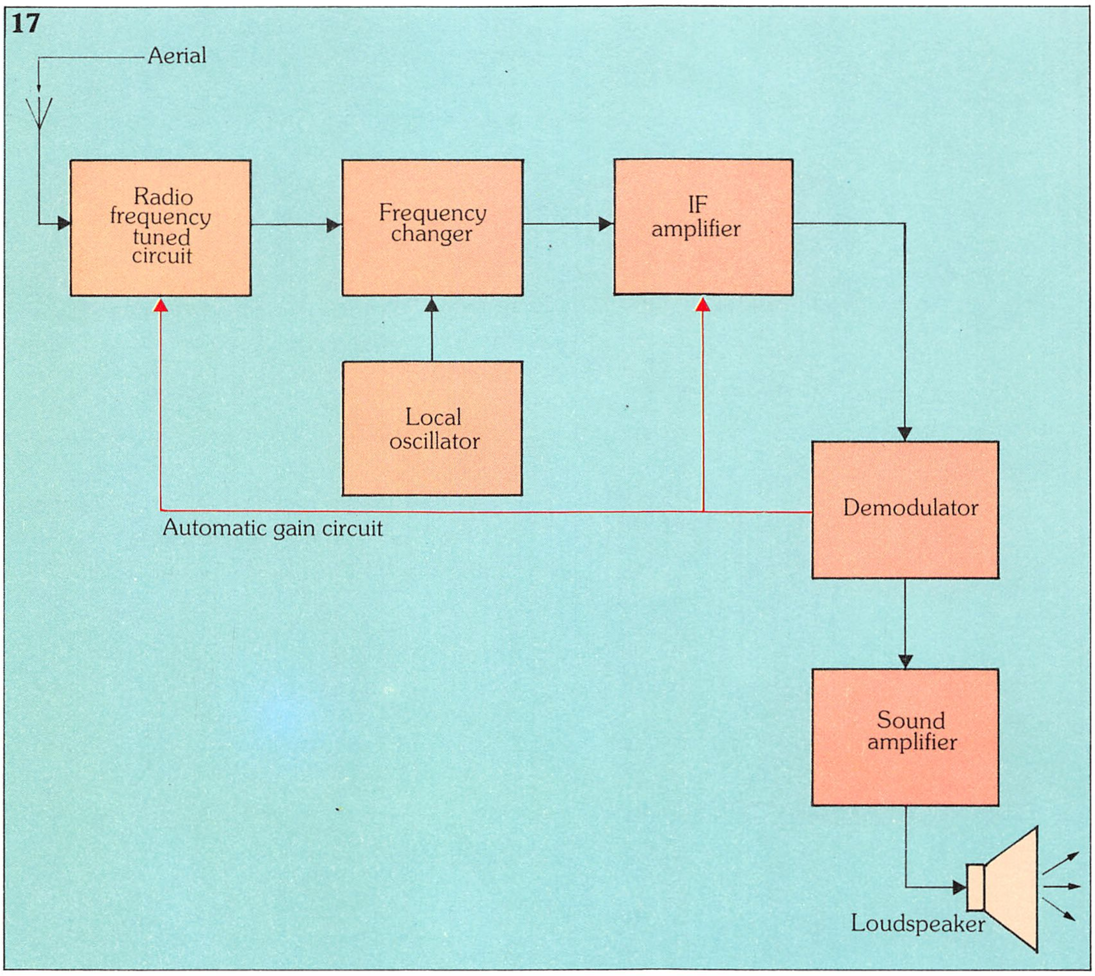


16. Typical signals occurring at points A, B, C and D of the AM radio receiver shown in *figure 15*.

amplified and applied to a loudspeaker.

Typical signals occurring at points A, B, C and D of the AM radio receiver shown in *figure 15* are shown in *figure 16*. Signals in an FM radio receiver (monaural) would be similar although those at points A, C and D would obviously be frequency modulated and not amplitude modulated.

The production of a difference frequency at point C by the application of a radio frequency signal and a local oscillator signal to the frequency changer is known as **beating** or **heterodyning**. The same effect is heard if two notes (just out of tune) are played simultaneously on a musical instrument. The resultant beats occur at the difference frequency between the two notes. So if, say, one note is played at 440



17. An automatic gain circuit (AGC) applied to a superhet receiver.

Hz, and another at 439 Hz, beats will be heard repeating at 1 Hz, i.e. one/sec.

A radio receiver operating on this heterodyning principle is sometimes known as a heterodyne receiver. It is usually, however, abbreviated and called a **superhet receiver**.

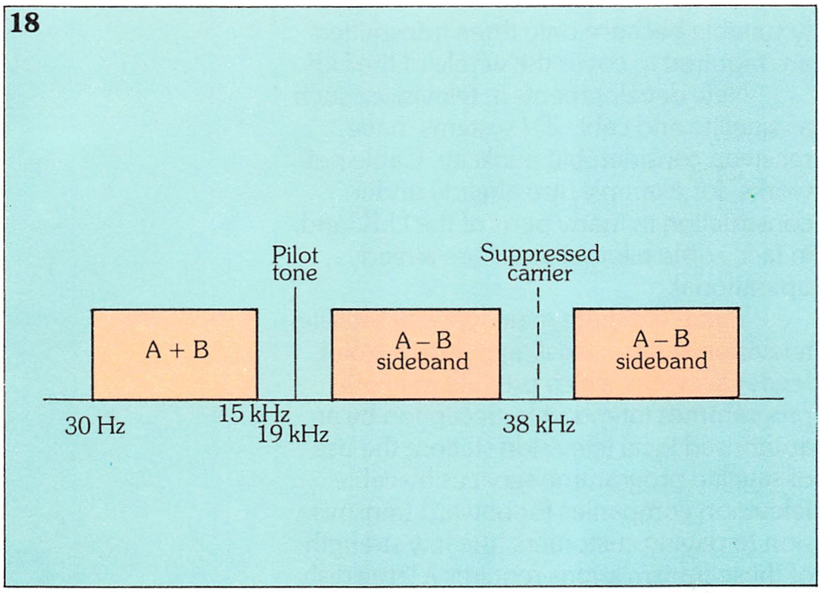
The basic superhet receiver of *figure 15* has a number of failings. For example, the audio output power depends very much on the strength of the receiver signals from the aerial. If these signals are low in strength the audio output power may also be low, but if they are high in strength the audio power may be so high they may be distorted. This distortion can be overcome by including an automatic gain circuit (AGC) in the receiver which automatically adjusts the overall gains of the amplifiers within the receiver depending on the signal strength, thus maintaining a more or less constant strength audio signal. *Figure 17* illustrates how AGC may be applied to a superhet receiver where a voltage proportional to the magnitude of the sound output from the demodulator is used to alter the gain of the radio frequency and intermediate frequency amplifiers.

Stereo radio

The radio receivers we have seen so far are specifically monaural, i.e. they receive and demodulate a single channel of audio message signal. Certain FM broadcasts within the U.K. are stereo, i.e. two channels.

The simplest method of broadcasting stereo message signals would be to use two separate radio channels – one to transmit the left hand (which we can call, say, A), the other to transmit the right hand (say, B) stereo signals. The BBC's first experimental stereo broadcast in 1926, in fact, used two transmitters but such a method is obviously wasteful and unsatisfactory, not least because the mono-listener can only receive one of the two signals giving an unbalanced result.

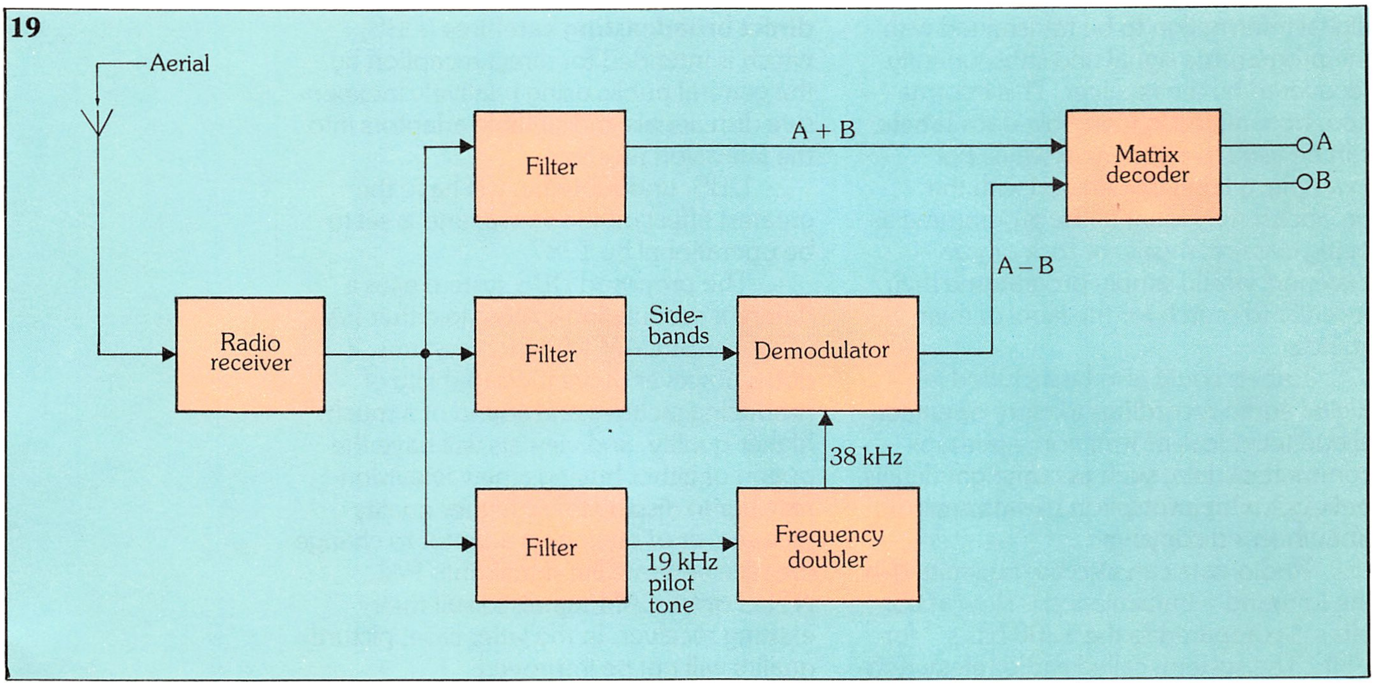
Things have changed considerably



difference between the two (A – B). The sum signal is similar to the signal derived from a single microphone and therefore provides a programme source for mono receivers.

Figure 18 shows a spectrum of a transmitted stereo multiplexed message signal and we can see that a **pilot tone** of 19 kHz frequency is transmitted alongside the sum signal. The difference signal is frequency modulated around a suppressed 38 kHz carrier in two sidebands. This whole spectrum is transmitted as a message signal using a carrier in the range 88 to 98 MHz.

The stereo receiver, a block diagram of which is shown in *figure 19*, incorporates a **matrix decoder** – a circuit which



18. A spectrum of a transmitted stereo multiplexed message signal showing that a 19 kHz frequency pilot tone is transmitted alongside the sum signal.

19. Block diagram of a stereo receiver.

since 1926 of course, and a system has been adopted worldwide which allows high quality reception of broadcast stereo transmissions. The system is fully compatible with monaural radio receivers, i.e. they can receive signals as if they were monaurally transmitted. This system is a multiplexed system in that two full bandwidth message signals are combined and radiated from a single transmitter within the frequency channel normally allotted to one monophonic programme. One of these signals comprises the sum of the two signals (i.e. A + B) the other comprises the

effectively adds the sum and difference signals to obtain the left-hand signal:

$$(A + B) + (A - B) = 2A$$

and subtracts the sum and difference:

$$(A + B) - (A - B) = 2B$$

to obtain the right-hand signal.

To demodulate the difference signal, the 38 kHz carrier (suppressed at the transmitter) must first be regenerated. This is simply performed by doubling the received 19 kHz pilot tone frequency.

The radio receiver block of *figure 19* is a conventional superhet FM receiver.

Conclusion

As we have seen, the principles of broadcast radio and television are similar. Both rely on the transmission of message signals, by modulation of carrier signals, over the air waves. The superhet receiver principle is common to both television and radio receivers and has served a useful purpose in these applications for many years now.

But what of the future? What can we expect to be commonplace in the world of television and radio communications over the next few years?

Radio data

One of the latest developments in radio communications is a data system for use with FM broadcasts. The system enables digital information to be transmitted with the programme signal and subsequently recovered by the receiver. This information, transmitted in inaudible **data labels**, can be used in a variety of ways. For example, a label transmitted with the broadcast may identify the programme as being classical music, or rock or jazz. Listeners would simply programme their receiver to search for the label of their choice.

Labels could also be included in digital audio recordings to carry operational and technical information – perhaps commercial data, such as copyright details, or keys for the protection of data against unauthorised copying.

Radio data can also be transmitted in the LF band – transmission is slow at 25 bits s^{-1} compared to the 1200 bits s^{-1} for VHF. The system, called **radio teleswitching**, superimposes an inaudible data signal onto the main carrier by modulating the carrier's phase. It can be effective in areas such as basements or steel framed buildings because of the low frequency nature of the transmission.

Radio teleswitching is used by the electricity supply industry to control electricity meters in private homes. It enables fine control of customers' tariff controlled appliances, such as storage or water heaters, and so if used nationally, would enable the industry to balance its power station load more accurately and efficiently. The 200 kHz frequencies are particularly suitable because only three transmitters are required to cover the whole of the U.K.

New developments in television, such as satellite and cable TV systems, have received considerable publicity. Cable networks, for example, are already under construction in many parts of the U.K. and, in fact, some pilot schemes are already operational.

There are three main types of satellite television service: existing **point-to-point feeder services** which provide network programmes intended for reception by an authorised local television station; the use of satellite programme services by cable television companies for onward transmission to paying customers (the low strength of these transmissions requires a large dish aerial to capture picture quality); and **direct broadcasting satellites** (DBS), which is intended for direct reception by the general public using relatively inexpensive dish aerials and suitable adaptors into the television receiver.

DBS, undoubtedly, will have the greatest effect on the viewer and is set to be operational by 1987.

The proposed DBS system uses a different modulation system to either PAL, NTSC or SECAM television systems. It does, however, have the capability of producing pictures and sound of a much higher quality, and viewers will have the option of either buying a new television receiver to display these higher quality pictures, or of buying an adaptor to change the received satellite signals into PAL, NTSC or SECAM signals to suit their existing receiver. In the latter case, picture quality will not be improved.

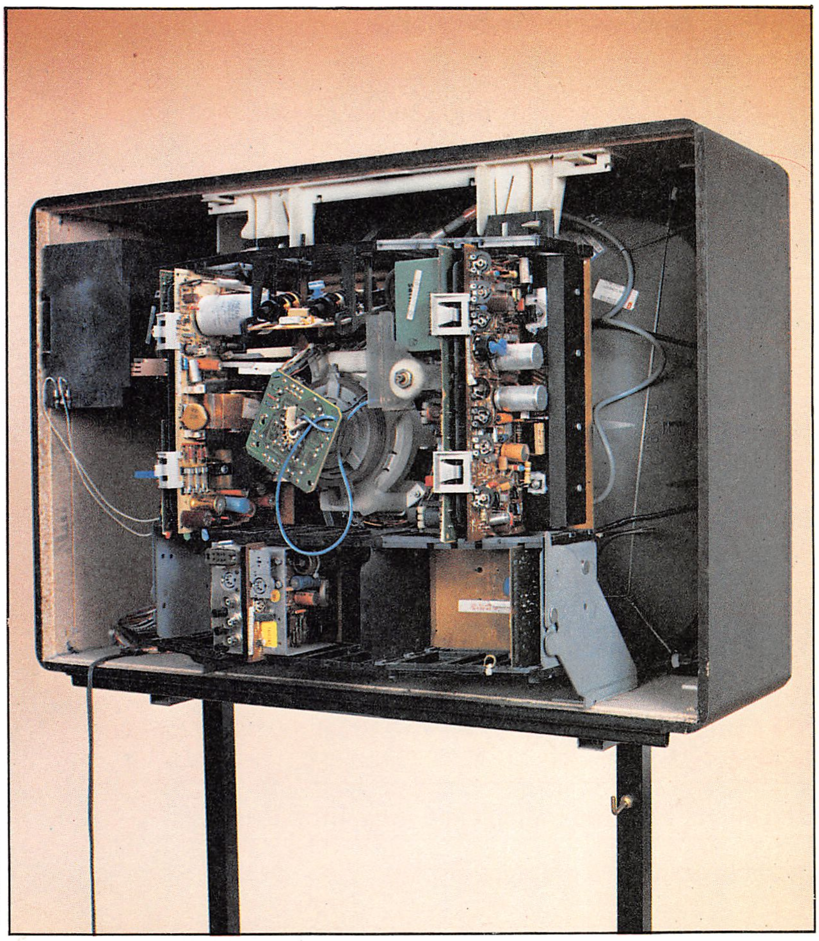
Satellite and cable television systems will be covered in greater detail later.

Teletext

Television receivers may be used for the reception of information other than broadcast pictures, of course. **Teletext** is the generic name for broadcasting systems in which pages of text and graphical information are transmitted to the viewer; viewers can also interrogate a library of text information. Teletext's main advantage over other media is speed - it has no copy deadlines and news can be broadcast within minutes. Teletext libraries also in-

clude social and business information.

Teletext information is transmitted in coded form on otherwise unused television lines and the decoded pages can be displayed in place of, or added to, the television picture. In the U.K., the BBC and ITV 625 line transmissions carry teletext information services known respectively as **Ceefax** and **Oracle** on television lines 17, 18, 330 and 331. These are transmitted in the field blanking interval of the television signal. Normally, these lines are inactive lines, out of view, although they will be seen if the set has been incorrectly adjusted.



Neville Miles

Above: inside a colour television set.

Each line of pulses contains the coded data for one row of standard size text or graphics rectangles. The rows for a given page are transmitted one after the other until the page is complete and then the next page is transmitted and so on. It is rather like a conveyor belt with coded rows of text being loaded one after the other.

If every line of a page had text or graphics, the page would take 0.24 seconds to transmit with interlacing – the average access time is seven seconds.

The binary pulses on the data lines are grouped into two sections: the first establishes which row of text it is on the page; and the second defines the contents of that row.

Each letter, number or graphic symbol is represented by an eight digit binary number. For the first row of the page, the first section also establishes which page is coming up and its time code.

Additional circuits are needed in the receiver to decode teletext. First, data extraction and recognition circuits examine the incoming conveyor belt of data lines and extract those signals which make up the selected page. The data from these lines is stored usually in RAM so that the page can be displayed at the same rate as the normal picture. The binary number codes are then translated to their corresponding characters or graphics patterns in ROM. Finally, a video scan of the page is switched onto the screen.

Teletext services can either be received using a receiver equipped with an integral teletext decoder or on a standard receiver used in conjunction with a separate teletext adaptor.

The technical specifications of Ceefax and Oracle are identical. The decoded pages are intended to be displayed in place of, or in addition to, the television and this allows the insetting of newsflashes or subtitles in the picture.

There is another use for television receivers in which data may be displayed on screen. **Viewdata** services are transmitted to the viewer's television set over telephone lines; the main U.K. example of viewdata is BT's Prestel. The information base is wider, and the system can be interactively used by the viewer in conjunction with a special keyboard to transmit his own requirements to a third party on the system.

The latest innovation is the transmission of **telesoftware** where Ceefax pages are used to carry computer programs which can be downloaded directly into home computers.

Teletext and viewdata systems will be further discussed in a later issue.

Glossary

adjacent channel interference	interference, caused within a radio receiver, from reception of unwanted radio transmissions
blanking level	voltage level of a composite video signal which causes the display to be black, and blanked
colour burst	signal transmitted in the video signal to enable the television receiver to distinguish between received colour difference signals
colour-difference, chrominance signals	R-Y, the difference between the red signal and the luminance signal; and B-Y, the difference between the blue signal and the luminance signal
dichroic mirror	mirrors which reflect light of a single colour while passing light of all other colours
field blanking lines	25 inactive (i.e. not displayed) lines between each field
frame	two fields
heterodyning, beating	principle of operation of the superhet radio receiver, in which the combination of two signals produces a third signal, the frequency of which is the difference between the combined signal frequencies
intermediate frequency	the third signal frequency, produced by heterodyning two radio frequency signals in a radio receiver
local oscillator	oscillator in a radio receiver which is combined with the received signal to produce the intermediate frequency
luminance signal	monochrome signal relating to brightness. Is formed by a combination of 30% of the red signal, 59% of the green signal and 11% of the blue signal
negative modulation	modulation of a composite video signal where a greater voltage causes a decrease in displayed brightness
pilot tone	tone transmitted in an FM stereo multiplexed radio broadcast
pixel, picture element, or pel	smallest addressable point on a graphics display
sub-carrier	used to transmit chrominance signals on top of luminance signals
superhet	heterodyne. A principle used in radio receivers in which a received signal is mixed with a local oscillator's signal to produce an intermediate frequency signal
synchronising pulses	pulses in a composite video signal which initiate both line and field scans in a television receiver
vestigial sideband modulation	form of envelope amplitude modulation in which only part of one sideband is transmitted. The other sideband is transmitted whole

ELECTRICAL TECHNOLOGY

Three-phase distribution

The different ways in which loads can be connected to a three-phase distribution system are shown in *figures 1b*, *c* and *d*. We will assume this system is working at 440 V (remember, this means that the **line to line** voltage is 440 V).

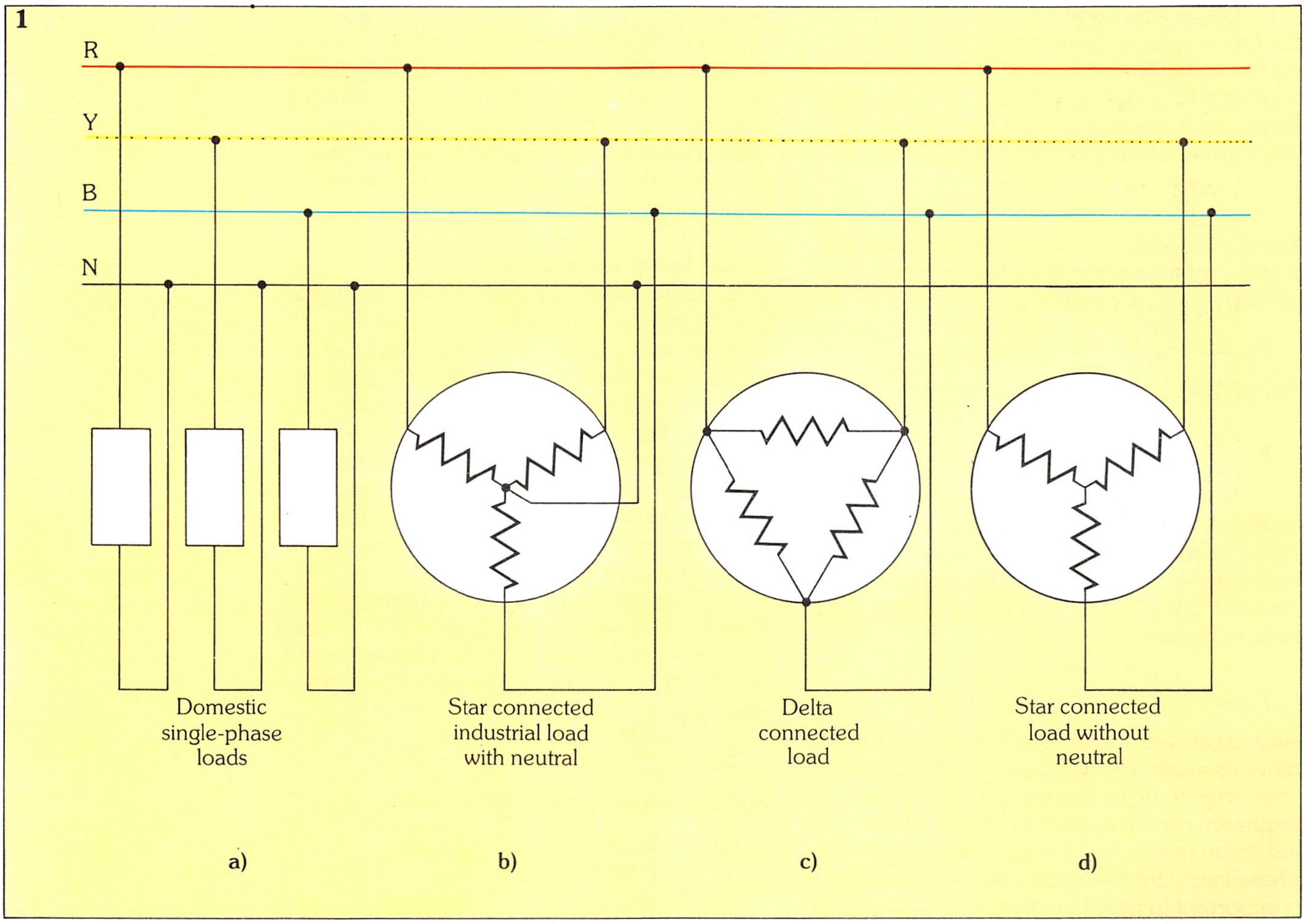
A number of single-phase loads are shown in *figure 1a* – these would usually be domestic consumers, connected between one of the phases and the neutral line. The voltage available to these consumers is the phase voltage, V_P, which is equal to $V_L/\sqrt{3} = 250$ V (approximately). Domestic installations in the U.K. are normally supplied from only one phase, so as to keep the voltage at a reasonably safe level. However, this is not the case in some other countries.

We know that the load on the three phases has to be kept as balanced as possible, thus in each street, one house is connected to the red phase, the next to the yellow, the third to the blue, and so on. The currents in the various phases of the generator will then become approximately equal all the time.

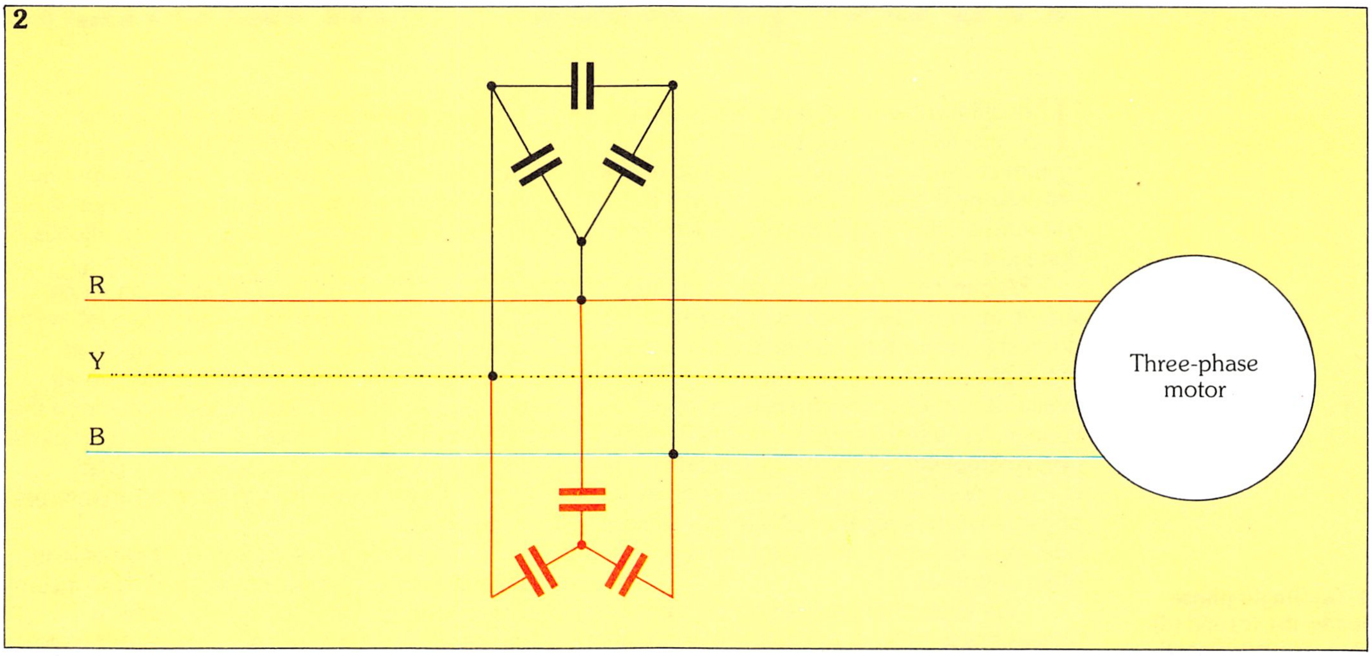
The load placed on the electricity supply by the domestic consumer is generally purely resistive: for example, heaters, cookers and lights. Only a small part of the power taken is reactive, i.e. that required by an appliance using an electric motor, and although some domestic motors may take one or two kilowatts of power, they are used for relatively short periods of time. This means that the overall average effect of the load's power factor on the generator is small. Fluorescent lights also take a lagging current from the supply, but their power factor is usually corrected by an internal capacitor.

The three-phase motor is a type of load restricted to industrial and commercial consumers. These can either be synchronous or inductive and (as we shall see in a subsequent *Basic Theory Refresher*) will take a current from the supply that contains a reactive component. This will either lag or lead, depending on the type of the motor. Some consumers also

1. (a) Single-phase loads; (b), (c) and **(d)** show the different ways in which loads can be connected to a three-phase distribution system.



2



2. Three capacitors connected in delta.

use large electric furnaces connected to the three-phase supply and these take currents which are entirely in phase with the supply voltage.

Let us now consider a three-phase electric furnace, comprising three equal heating elements, each of resistance R. If these are connected in a star, as in *figure 1d*, the voltage across each element is the phase voltage, V_P. The current flowing in each resistance is:

$$I_P = \frac{V_P}{R}$$

The power generated in each element is V_P^2/R and the total power in the whole furnace will be:

$$P_{star} = \frac{3V_P^2}{R}$$

Since $V_P = V_L/\sqrt{3}$, the power is:

$$P_{star} = \frac{V_L^2}{R}$$

Now consider the same heating elements arranged in a delta (*figure 1c*). The voltage across each one is V_L, and the current is therefore V_L/R. This leads to the power in each load being V_L^2/R. The total power in the whole furnace is thus:

$$P_{delta} = \frac{3V_L^2}{R}$$

As you can see, for an identical system, three times as much power is developed when it is connected in delta, than in star. The same argument can be applied to a three-phase induction motor, and for this reason, three-phase loads are invariably connected in delta. It is important to note here that, when connected in delta, larger currents flow, so the cross-sectional area of the conductors used needs to be correspondingly larger. The insulation also needs to be thicker as greater voltages are involved.

The smaller current flow in star connections is usefully employed when starting three-phase motors. When switched on, the motor is connected in star, and it is switched over to delta connection as its speed increases. This minimizes the very large current flow when starting any motor.

Power factor correction

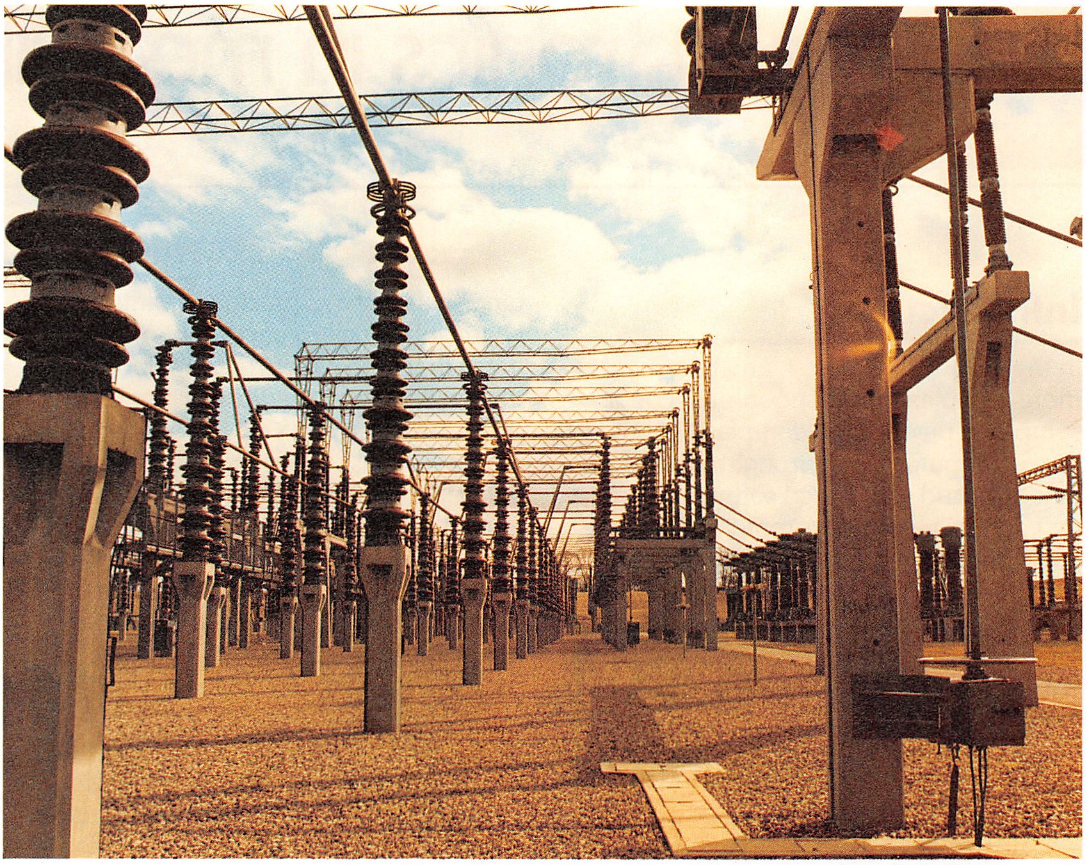
In an earlier *Basic Theory Refresher* on single-phase systems, we saw that the current is minimum for a fixed power delivered to an installation, when the power factor is unity – this is still true in the case of three-phase systems. The power factor can be improved by connecting three capacitors – either in star or delta – across the load terminals, to compensate for a lagging current caused by a three-phase motor. The size of the capacitors can be determined in a similar way as before.

Consider a load that is consuming power P watts at a power factor of cos ϕ and we want to improve the power factor to a value of cos θ. We find that the apparent power of the load is given by:

$$S_L = \frac{P}{\cos\phi}$$

and the reactive power of the load is:

$$Q_L = S_L \sin\phi = \frac{P \sin\phi}{\cos\phi}$$



The Research House/CEGB

Right: the CEGB's speciallly designed Mark II 400 kV Wymondley substation in Hertfordshire. Self-supporting aluminium lattice structures in the form of a bridge are used in place of the massive steel gantries usually seen.

$$\phi = P \tan \phi$$

As we want to reduce the power factor to cos θ, with the same power supplied, we can see that the reactive power taken from the supply is:

$$Q_S = P \tan \theta$$

To obtain this improvement, we need to connect a set of capacitors that take a reactive power Q_C, which is given by:

$$Q_C = Q_L - Q_S$$

If these capacitors are connected in delta, as shown in black in *figure 2*, we can see that the current in each capacitor is $V_L \omega C_\Delta$. The total reactive power of the delta of capacitors is:

$$Q_C = 3V_L^2 \omega C_\Delta$$

so that:

$$C_\Delta = \frac{Q_C}{3V_L^2 \omega}$$

Alternatively, if the capacitors are connected in star (as shown in red), the voltage across each capacitor is $V_L/\sqrt{3}$ and the current is $V_L \omega C_S/\sqrt{3}$. The total reactive power in the three capacitors is:

$$Q_C = 3\left(\frac{V_L}{\sqrt{3}}\right)\left(\frac{V_L \omega C_S}{\sqrt{3}}\right)$$

$$= V_L^2 \omega C_S$$

So the value of the capacitors needed is:

$$Q_S = \frac{Q_C}{V_L^2 \omega}$$

The value of the capacitor needed for delta connection is a third of that required for star, so delta is usually the preferred connection. However, it must be remembered that capacitors in the delta configuration must withstand a higher voltage than those in the star.

It is important to realise that this method of power factor correction is only applicable to *balanced* three-phase systems. In the case of unbalanced loads connected between individual lines and neutral, power factor correction must be carried out on each individual phase, treating it as a separate single-phase load. □

7

COMPUTERS & SOCIETY

Computers and electronics in medicine

Introduction

In the industrialized west, any treatment more complex than a spoonful of cough-syrup or a piece of sticking-plaster could see a computer being brought into use. Finances and specialised software permitting, the utilisation of computing power could be an everyday occurrence in the pharamacy, checking stock levels and dosages; in medical practice administration, scheduling appointments and maintaining patient records; in the clinics, analysing data from ECG and EEG recordings to blood pressure measurements, and so on; and in almost every other aspect of a modern health service.

In medicine, as in so many other fields, computers were first applied to the repetitive processing of large amounts of data, though the move towards a statistical basis for the practise of medicine began a decade or so before 'the computer age'. Today, without the computer, the processing of the data collected in, for instance, the trial of a new drug, would be quite impractical.

The varied applications which have been found for computers in the medical field go far beyond mere statistical processing, however. A large general hospital may have upwards of a thousand in-patient beds, and be handling several thousand out-patients in up to twenty or more different clinics each week. The management of an enterprise this big presents some severe (and unique) logistical problems. Computers have now become indispensible in hospital management: monitoring bed and clinic allocations; keeping patient and staff records; budget and inventory control; and forward planning.

These management and statistical functions require large mainframe computers and the expense of buying and running such systems precluded, until recently, the introduction of computing into other areas of health care provision. However, the availability of cheap and increasingly powerful microprocessor based instrumentation has shifted the emphasis of computer utilisation from statistical analysis and large file handling jobs to diagnostic and clinical care applications.

Microprocessor based instrumentation is increasingly finding its way into direct patient care facilities. The electrical signals associated with the body generally range from DC to a few hundred hertz, and rarely exceed a few kilohertz. Analogue-to-digital conversion of signals in this range is quite straightforward, and the subsequent digital processing of the signals yields considerably more information than previous analogue techniques (as well as presenting the information in a more easily interpreted format).

Interactive diagnostic systems (expert systems) are also being developed. These are not yet at the stage where doctors could be made redundant but their greater body of medical knowledge will also significantly improve the speed and accuracy of diagnoses, especially in more difficult cases.

The impact of computers on many areas of the health service has undoubtedly had many benefits and it is worth noting that there is one very important field of medical science which would not exist at all without them. **Medical imaging**, which includes such techniques as computerised tomography (C.T. scanning) and ultrasound scanning, as well as new developments in magnetic resonance imaging (nuclear magnetic resonance), requires the processing of huge quantities of data. The pictures of sections of the body, head and internal organs produced by these techniques would be impossible without computers to manipulate the raw data from the scanning machines.

Medicine and statistics

The importance of the statistical analysis of data collected in every area of the health service cannot be understated. Until the early part of this century, medicine was practised along largely empirical lines. That is to say, upon finding a treatment successful in one case, the physician or surgeon would repeat the same treatment on all patients showing the same or similar symptoms. Little or no account was taken of other factors relating to the patient – only the immediate indications of the problem were used in diagnosis. This meant (amongst other things) that chronic disorders and diseases relating to specific geographical or socio-economic conditions often went unrecognised as such, and indeed under these conditions they became untreatable.

The introduction of large mainframe computers with powerful statistical analysis programs enabled information about the incidence of disease to be correlated with other information about the subjects. Thus, for instance, by examining the statistics relating to lung cancer, it has been shown that cigarette smokers are about five times as likely to die from cancer as non-smokers.

Statistical packages used in the U.K. include such programs as: BMD (BioMeDical computer package); SPSS (Statistical Package for Social Sciences); GENSTAT; and GLIM. Each have characteristics which make them more suitable for particular jobs. GLIM, for instance, is designed to carry out transformations and can run powerful curve-fitting routines. The strength of SPSS lies in its ability to sort and extract subsets from large data files and it has over 30 options in various methods of statistical analysis. Such powerful programs are necessarily very complex and are usually run by professional programmers or statisticians attached to the staff of a hospital or research institute, rather than medical or research staff themselves.

The use of statistics is prevalent in all branches of medicine and allied disciplines. Much of the work being done, as exemplified by studies of the effects of smoking or the causes of coronary failure, is aimed at preventing illness rather than curing it once it has occurred. As research continues to highlight new information pointing the way towards better health, it is important to remember that this information should then be disseminated on a large scale – here, the activities of health education centres become very important.

Right: portable electocardiograph

The Research House/Albery Instruments

Medical services management

Until recently, only the management of larger hospitals and higher levels of administration in the health service could expect to call on assistance from computers. Yet the amount of form filling and quantities of data collected in even a small general practice can be mountainous. A couple of examples will demonstrate this.

In the U.K., a dentist has to complete and return a form FP17 to the Dental Estimates Board each time a patient is treated under the NHS. The form must be completed, no matter how trivial the job, or the dentist will not be paid. It has recently been estimated that filling in these forms can take up as much as four weeks of the dentist's time in one year. And besides this, the dentist is obliged to keep detailed records on all patients for at least two years following their last visit. A dentist may have a patient list of 5000+, and be carrying out 250 consultations a week.

A doctor in general practice has similar problems. A general practitioner receives over two million words of information about drugs each year. A typical family doctor sees around 5000 patients in a year, carrying out between 10 and 15,000 consultations. In 1972, one doctor wrote that he had recorded 31,110 symptoms and advised 18,243 treatments in the previous year. Extrapolating these figures to the U.K. as a whole, he estimated that approximately half a billion (500,000,000) symptoms may be recorded in a year; the number of prescriptions issued would be of the same order.

Given these figures, it might seem obvious that primary health care is ripe for computerisation – however, there are obstacles. The sheer bulk of data being handled requires a considerable investment in hardware – although as computers become cheaper and more powerful this is likely to become less important. Perhaps more of a hindrance to their acceptability is the installation time needed for a system to be up and running in the surgery. The transfer of records from the current manual systems and learning to use the system effectively take up time which is generally in very short supply in a general practice.

Another, and perhaps more obvious obstacle is that of gaining acceptance for new technology in a profession proud of its traditions of personal attention and professional secrecy. Many doctors view computers with more than a hint of suspicion, worried by stories of computer systems being broken into by other computer owners gaining access via the public telephone network. Furthermore, until recently there was very little software actually designed for this market – that which was available was often a poor adaptation of software written for small businesses and consequently of little use in the unique environment of a doctor's or dentist's practice.

These obstacles are now being overcome. Various projects, under way in the U.K., are seeking to develop standards to enable the ready transfer of data between surgeries, regional health authorities, and central governmental bodies. One such

Below: the LSC 7000 ultrasound linear/sector scanner.

The Research House/Picker International

project, in Exeter in south-west England, has moved away from the concept of a mini or microcomputer in every surgery. Four health centres, with fifty terminals between them, are linked to a central computer handling all medical records. Local hospitals also have access to the system for out-patient and in-patient references, and for reporting back on X-rays or laboratory analyses requested by a GP.

Each patient's record summarises medical history along with a full list of treatments – drugs prescribed, dosages, length of course of treatment. Notes on allergies and adverse reactions, and an attendance record (surgery, home visits, consultations) are also included. Back-up records are kept on microfiche, which can also be consulted on home visits with the aid of a hand-held microfiche reader.

Confidentiality of records on a system like this is accomplished by the allocation of passwords to authorised users: different passwords limit the range of records available, or even the information released – surveying the incidences of a particular disease, for example, does not require the patient's name to be released. The Exeter project has been studied by a working party from the Royal College of General Practitioners, who concluded that a higher level of confidentiality can be achieved on a computerised records system than is possible under the current manual system.

In the U.S. there is a greater acceptability of computers in all types of small consultancy. Paradoxically, the high degree of centralisation typical of European health services has probably hindered the introduction of computer systems at the primary level. End-users have waited for the central authority (e.g. the NHS in the U.K.) to provide some sort of lead, or to set standards. The lack of a central regulatory body of this type in the 'free-market' health care system of the U.S. has allowed doctors to acquire any computer system they may feel to be adequate for their own personal needs. Of course, the free market system also provides a further incentive to computerisation. Increased efficiency brought about by the use of computers in scheduling appointments and preparing bills has tangible results in the form of higher profits.

New technology and patient care

The fact that living organisms respond to electrical stimuli was first demonstrated more than a century ago. Later, with the invention of the cathode ray tube, physiologists found in the oscilloscope an instrument sensitive enough to measure the electrical signals produced in a living body. Since that time, the disciplines of electronic engineering and medical and biological science have co-operated to produce ever more complex instruments for diagnosis, treatment and research.

The range of instruments used directly in the treatment of medical conditions is quite small compared with the wider range of equipment utilised in patient monitoring. Typical of these machines is the **radio-frequency diathermy** equipment used in operating theatres. A miniature probe is used to apply a high power radio-frequency signal to the edges of the incisions made by the surgeon. This effectively cauterizes the wound, sealing blood vessels and killing surface tissue, thereby preventing infection.

But it is in the areas of diagnostics and patient monitoring that the microprocessor has had its greatest impact. Instruments used as diagnostic aids range from relatively simple machines, such as cardiac rate-meters (which time the interval between successive heartbeats measured from an electrical signal picked up at some point on the chest), to complex computerised tomography X-ray scanners.

Many, if not most, of the instruments in use in diagnosis and monitoring measure electrical signals either directly from the patient's central nervous system (via skin surface electrodes) or from one of the wide range of transducers used to measure other physiological phenomena. However, there are two problems common to many of the more frequently used transducers. The first is that these transducers exhibit a very high impedance when used as voltage sources. High input impedance amplifiers therefore have to be used, with their associated problem of the pick-up of stray electromagnetic signals.

This problem is highlighted in an EEG

(ElectroEncephaloGraph) which records the brain's voltage waves (it is also apparent in ECG – ElectroCardioGraph – recordings). The solution to the problem is to use head mounted preamplifiers. Electrodes are positioned in various locations on the scalp – on top of an electrically conductive gel.

However, the brain is protected from the hard bone of the skull by a conductive fluid which tends to disperse the signals which the EEG is attempting to measure, and the bone itself is an insulator which attenuates the signals. Because of this, the signal at the scalp has a typical amplitude of about 20 μV (although this can peak at 100 μV under certain circumstances) and the impedance of the electrode/scalp circuit is usually 1 to 3 kΩ.

Under these conditions the most common problem is interference in the amplifier and so all possible interference sources must be eliminated before recording. One advantage of the EEG is that the signals of interest generally occur in a frequency band no higher than 25 or 30 Hz.

The second problem is **drift** – that is, the varying of an electrical response to a given stimulus with time. With pH transducers, for instance, the only way of obtaining meaningful measurements is to regularly recalibrate the associated amplifier after comparing the transducer's response to a substance of known pH value. For some transducers, however, it is possible to overcome the problem by utilising complex analogue circuitry. Digital processing or the application of corrective software algorithms in a microprocessor system may also help. Self-calibrating systems represent one of the more important advances in patient monitoring brought about by the microprocessor. Reduced maintenance, shorter calibration time and greater reliability all contribute to better and more efficient patient care facilities.

A further advantage of microprocessor based equipment is that it can present information in a much more easily interpreted format. For instance, before the micro, a patient in a cardiac acute care ward could be attached to a cardiac ratemeter, a pacemaker, a monitor (to display the ECG waveform) and/or an ECG chart recorder, as well as blood pressure monitors and perhaps other equipment thought necessary by medical staff. This equipment needed to be constantly monitored, and perhaps reams of paper carrying the previous 24 hours' ECG recordings would need to be studied to detect any irregularities.

By contrast, typical microprocessor equipment now in use can display: a picture of the ongoing ECG; a numerical display of cardiac rate and blood pressure; and also a graphical representation of rate with upper and lower alarm limits – all on a single screen. Many modern cardiac monitoring systems are capable of recognising anomalous ECGs and they may be programmed to sound an alarm when an anomaly is recorded, or, in less acute cases, to store data relating to a period of time either side of the event. Data recorded in this way can then be recalled for further analysis at a later date.

The complete analysis of biological signals and subsequent diagnosis by com-

Above: NMR consoles and operators. (Photo: Picker International).

puters is still a long way off, partly because the physician bases his final diagnosis on more than just a list of symptoms – personal experience and knowledge of the patient's background play an important part in his decision.

Expert systems in medicine

The knowledge base in medicine, as in other professions, has fast expanded in recent years and continues to do so at an increasing rate – about 5% per year. It has been estimated that the current body of medical knowledge represents about ten times the amount of information that the human brain can handle. In the 1930s, American medical colleges realised that doctors could no longer be expected to retain a comprehensive knowledge of all symptoms and treatments. This problem was solved by the creation of specialists – some doctors became experts in child health, some in gynaecology etc. and there are now over one hundred recognised medical specialities.

General practitioners, however, are still expected to recognise any symptoms with which they may be presented – at least well enough to refer the patient to the relevant consultant for further investigation. For this reason, much effort is now going into the development of expert systems for general practice.

One essential prerequisite for such a system is the identification of the clinical facts relating indicators of disease to the actual complaint; this information, of course, can be easily stored in a computer system. Of equal importance, is the necessity to re-examine the education and training of prospective GPs. The amount of time currently spent memorising information which could more easily be committed to reliable data storage facilities will become unnecessary. What can then assume prime importance will be the human aspect of communication – getting the required information from the patient to form the basis for consultation with an information retrieval system.

One very comprehensive system embodying these principles, and performing some of the administration functions that we discussed previously, has been developed and is in regular use in a New York practice. The system has been named PRAKTICE, an acronym for 'Physicians' Records And Knowledge yielding Total Information Consulted Electronically'.

Information relevant to reaching a diagnosis is divided into four areas: patient history; clinical experience of similar cases; information selected from textbooks and thought to be pertinent; and information on research activities. PRAKTICE is designed in such a way as to allow **information nodes** to point to other relevant nodes, containing further classifications or other, related information.

As an example, a user consulting the node for gout could continue by moving to any of the nodes describing clinical signs, relevant diagnostic techniques, clinical courses, genetic information, etc. The nodes are connected by arcs enabling two-way movement in information searches, so that the user can just as easily start with symptoms and move towards a diagnosis. PRAKTICE includes nodes which indicate standard treatments, and because it also maintains patient records a quick decision can be reached by the physician on the most suitable treatment in a particular case. PRAKTICE also includes facilities for automatically billing patients.

However, it is likely to be some time before systems like this are accepted by a significant proportion of their potential users. Expert systems need large memory capabilities for data storage and this means that they are going to be very expensive – at current U.K. prices it is unlikely that the necessary hardware could be produced for less than around £10,000. There is also the inherent problem of conservatism in the profession, particularly if there is an implicit suggestion that the doctor's knowledge of medicine is not as complete as it ought to be. (See also *Computers & Society 6*.)

Medical imaging technology

Whilst medical imaging techniques include classical radiography using X-ray cameras, this discussion will concentrate on newer developments in which computers are used to construct images which would otherwise be impossible to generate. Currently, several methods are employed to acquire detailed images of cross-sections of

the body and its internal organs. Computerised **X-ray tomography** is probably the best known method, as it received considerable public attention in the mid-1970s, following the introduction of EMI's brain scanners. Other imaging methods in common use include **ultrasound scanning** and **nuclear medicine**.

Tomography is a radiological method that maintains a fixed distance between the X-ray source and the receiver, so that the ratio of their distances from a chosen plane in the body remains constant. This arrangement produces a defined image of that plane. When automated by a computer, the X-ray source and receiver are manipulated to produce a serial succession of scans providing a whole view of the brain or other organs.

Computerised X-ray tomography, or C.T. scanning, was first introduced in 1971 when the first brain scanner was installed at Atkinson Morley's Hospital in Wimbledon. It soon became apparent that this instrument, and its successors, would revolutionise investigation into brain disorders. Other techniques in use for brain investigations at the time were riskier, more invasive and involved much greater discomfort for the patient.

Early scanners were limited to brain studies because of the time taken to acquire data – typically in the region of five minutes. Holding the head still for this amount of time involves relatively little discomfort for the patient, however motion in the torso, from heart, lungs and intestines, made body imaging impossible until shorter scan times could be achieved. Nowadays scan times of between 2 and 15 seconds are common.

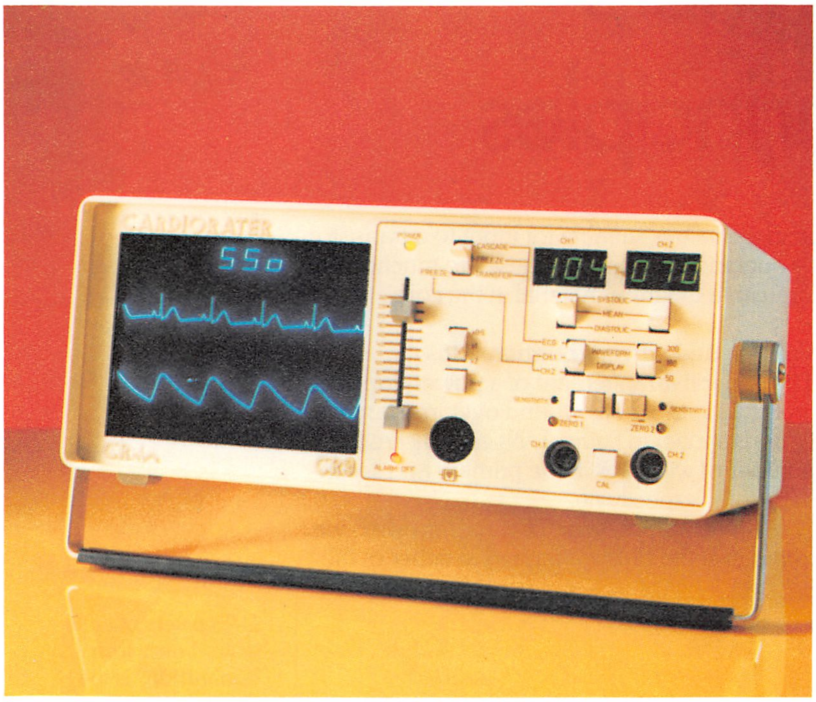
Scanners comprise an X-ray point source which is mounted opposite a detector: this arrangement passes either side of the patient. At a fixed point, the source emits a stream of X-rays. The strength of radiation detected is an indication of the density of material through which the radiation has passed – this depends on the quantity of bone, flesh, etc. and also on the relative absorption properties of the different components.

Once the absorption along one line through the subject has been measured, the source and detector move along a small distance and repeat the measurement along another, parallel line. This operation is repeated until the entire body has been traversed. The gantry on which the source and detector are mounted is then rotated through an angle of one degree and the traversing operation repeated; the complete scan is finished when the gantry has rotated through 180°. The data collected at the measurement points is entered into a computer where it is manipulated as a number of simultaneous equations, the solution of which provides absorption intensities for each pixel in the final image. The amount of data, and the processing power needed, means that typically a computer the size of a small mainframe is used – several modern scanners use the PDP-11 series can address up to 256 kbytes of RAM, and up to 4 Mbytes in an expanded version.

Although producing excellent image quality, early scanners were very slow. The second generation of scanners reduced scanning times to 18 to 20 seconds by employing multiple detectors opposite the X-ray source. This meant that at each point in the traverse, the scanner could measure a 10° fan of radiation. The gantry was then rotated through another 10° before the next traverse. Although scan times were much reduced in these systems, and body images could be created, they were still not fast enough to overcome all the problems of movement associated with internal organs.

Third and fourth generation scanners utilise a wide angle fan of radiation (typically 20° to 50°) and the source is rotated through 360°. Third generation machines still rely on an array of detectors mounted on the rotating gantry opposite the X-ray source. In fourth generation machines, however, a 360° ring of detectors is mounted stationary outside the orbit of the source. These machines can achieve scan times as low as two seconds. One further advantage is that the data is measured twice in the 360° scan, thus improving reliability of the image.

C.T. scanning still has certain limitations. Chief among these is the danger involved in exposing human tissue to radiation, a problem of particular concern when studying a developing embryo or



The Research House/Cardiac Recorders Ltd

Above: CR9 cardiac recorder showing ECG waveforms.

foetus. Because of this, ultrasound imaging is used.

Ultrasound imaging is a far more complex process than C.T. scanning in terms of information processing, but it is wholly non-invasive. Very high frequency sound waves are fired at the body by an emitter probe – the signals bounced back are then detected and these form the image – a process akin to sonar.

Nuclear medicine is a technique in common use for obtaining images of internal organs. A radioactive isotope is introduced into the body, at the location to be studied. The isotope is usually in solution, to be processed by the organ of interest. The image formed is in a form similar to an ordinary photograph, or X-ray image: two-dimensional showing the distribution of the isotope in the subject.

Nuclear magnetic resonance – future technology

The phenomenon of nuclear magnetic resonance, first discovered in the 1950s, holds out much hope for spectacular advances in research and clinical applications in both imaging and biochemistry/metabolic studies.

The term **resonance** refers to an effect which certain atomic nuclei exhibit under the influence of an extremely powerful magnetic field. These nuclei act rather like minute bar magnets, with definite North and South poles. They also spin on their axes – this spin is offset from the axis of the magnetic poles in much the same way as the Earth's rotational axis is offset from the magnetic pole. When a powerful external magnetic field is applied, the magnetic poles tend to change polarity, resonating between the old and new polarities at a characteristic frequency, emitting electromagnetic radiation as they do so.

It is this characteristic or resonant frequency, measured using resonant frequency pick-up coils, which provides the information. Typical resonant frequencies are around 270 MHz for the hydrogen isotope ^{1}H, and 73 MHz for the phosphor isotope ^{31}P. Several factors influence the resonant frequency, of which the most important is the molecular environment.

By magnetically exciting a sample and analysing the resultant resonant frequency output it is possible to obtain detailed information regarding the chemical composition of the sample. The importance of this for medicine is that instantaneous chemical analyses can be carried out on living, active material in cases which, with traditional techniques, would have destroyed the subject of the test. Thus the exact metabolic behaviour of a muscle can be studied under a variety of conditions, continuously, and if necessary over long periods of time. Magnetic resonance spectrometry, as this technique is referred to, has already overturned some long-held beliefs about the behaviour of chemical reactions in the body.

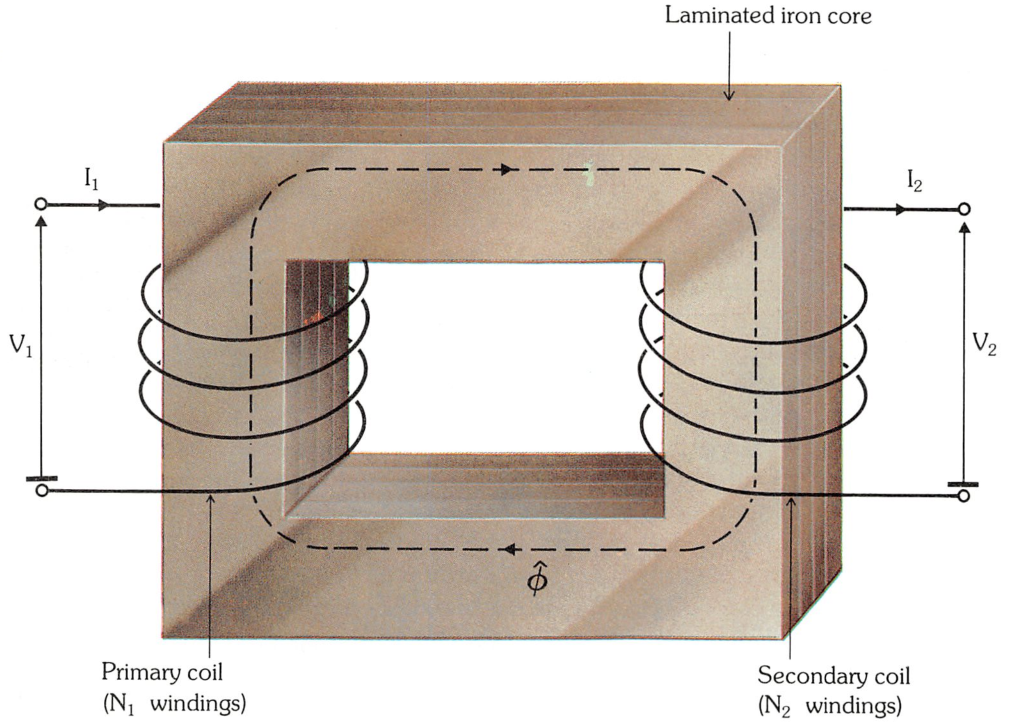
A second important application of NMR is that it provides a new method of imaging otherwise inaccessible parts of the body. Because the resonant frequency probe can be focused on any point in a 3-dimensional space, it is possible to build up pictures of the inside of the body which are beyond the capabilities of C.T. scanners. These images can reflect either the relative concentration of any injected isotope under study throughout a region by comparing the total power of the resonance signal at each point, or, because different molecules cause very small frequency shifts, they can be more selective and image specific chemicals.

ELECTRICAL TECHNOLOGY
Transformers - general principles

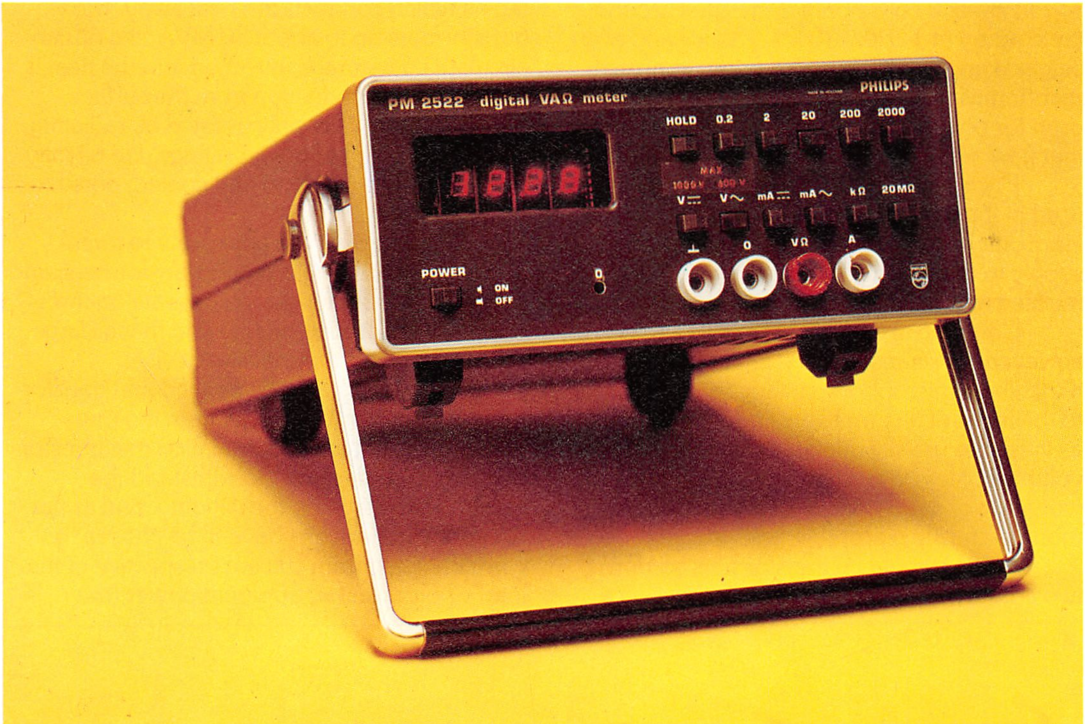
Transformers are used in practically all electrical circuits. They can either be employed to alter the voltage or current to an appropriate value or they may isolate one part of a circuit from another. For example, in supply systems, transformers enable power to be distributed at fairly high voltages over large distances and then transformed to a lower voltage for distribution to the consumer (such devices may stand 5 to 10 m high). At audio frequencies, transformers are often used to match the voltage and current provided by a microphone to those values suitable for input to a transistor or operational amplifier; these are much smaller, with a different material in the core.

Voltage and current relationships
The general arrangement of a power transformer, two sets of coils wound over iron laminations, is shown in *figure 1*. N_1 and N_2 are the number of windings in the primary and secondary coils respectively.

In an earlier *Basic Theory Refresher*, we found that the EMF generated in either of the

1


1. Two sets of coils wound over iron laminations is the general arrangement of a power transformer.



Right: Phillips PM 2522 digital multimeter.

windings is proportional to the rate of change of the flux linking with it. So, if the flux, ϕ, varies sinusoidally at a frequency f, with a maximum value $\hat{\phi}$, then the maximum value of the voltage induced in the secondary coil is given by:

$$\hat{V}_2 = N_2 2\pi f \hat{\phi}$$

The rms value of the secondary voltage is:

$$V_2 = \frac{\hat{V}}{\sqrt{2}} = 4.44 N_2 f \hat{\phi}$$

The voltage applied to the primary is equal and opposite to the induced EMF, so that:

$$V_1 = -4.44 N_2 f \hat{\phi}$$

From these two equations we can see that the ratio of the magnitude of these voltages is:

$$\frac{V_2}{V_1} = \frac{N_2}{N_1}$$

A **step-up** of voltage can be made if N_2 is larger than N_1; while a **step-down** results if N_2 is smaller than N_1.

As only a small amount of power is dissipated in a transformer, we can consider the power entering it to be equal to the power leaving it. If the load on the transformer is taken at unity power factor, we can therefore see that the power out, P_2, is given by:

$$P_2 = V_2 I_2$$

Similarly, the power entering the transformer, P_1, is:

$$P_1 = V_1 I_1$$

Since P_1 and P_2 are equal, we can see that:

$$\frac{I_2}{I_1} = \frac{V_1}{V_2} = \frac{N_1}{N_2}$$

This shows that the current flowing in the secondary of a voltage step-up transformer is smaller than that in the primary; the opposite is true for a step-down transformer.

We will now take a more detailed look at some of the applications of transformers that we have already outlined. Suppose that we wish to transfer power of 5 MW, at unity power factor, over a distance of 100 km, and supply it to consumers at 250 V. The current taken by the consumers is:

$$I = \frac{P}{V} = \frac{5 \times 10^6}{250} = 20\ \text{kA}$$

If the voltage at the consumer must be no more than 5% lower than that at the generator, then the resistance of the transmission lines needs to be kept sufficiently low. In reality, this specification would require the use of conductors with a cross-sectional diameter of about two metres, which is of course absurd.

If the same power is transmitted at 30 kV, then the current drawn will be only 167 A. The conductors would then need a cross-sectional diameter of about 25 mm – which is quite feasible. To achieve this, a step-down transformer is installed, close to the consumer, with a turns ratio of 120:1.

Power is commonly distributed around the country at 1.3 kV, 30 kV, 133 kV and even bigger voltages. By using step-down transformers in this way, the size of the conductors used for power transmission, and the currents that flow, are kept to acceptable proportions.

Power transformers are also commonly used in the provision of DC supplies to electronic circuits. They step-down the mains supply to the required value; the voltage is then rectified and smoothed.

Coupling at audio frequencies can be achieved with transformers, for example, matching a microphone or tape recorder head to the input of an amplifier. Transformers such as these are usually miniature devices (with volumes of less than 1 cm^3) with their coils wound over a pot-core, made from two moulded ferrite pieces.

Radio frequency transformers comprise two coils wound adjacent to each other on a cylindrical plastic former. A ferrite core is inserted into the centre of the former, and adjusting its position will modify the transformer's operating properties.

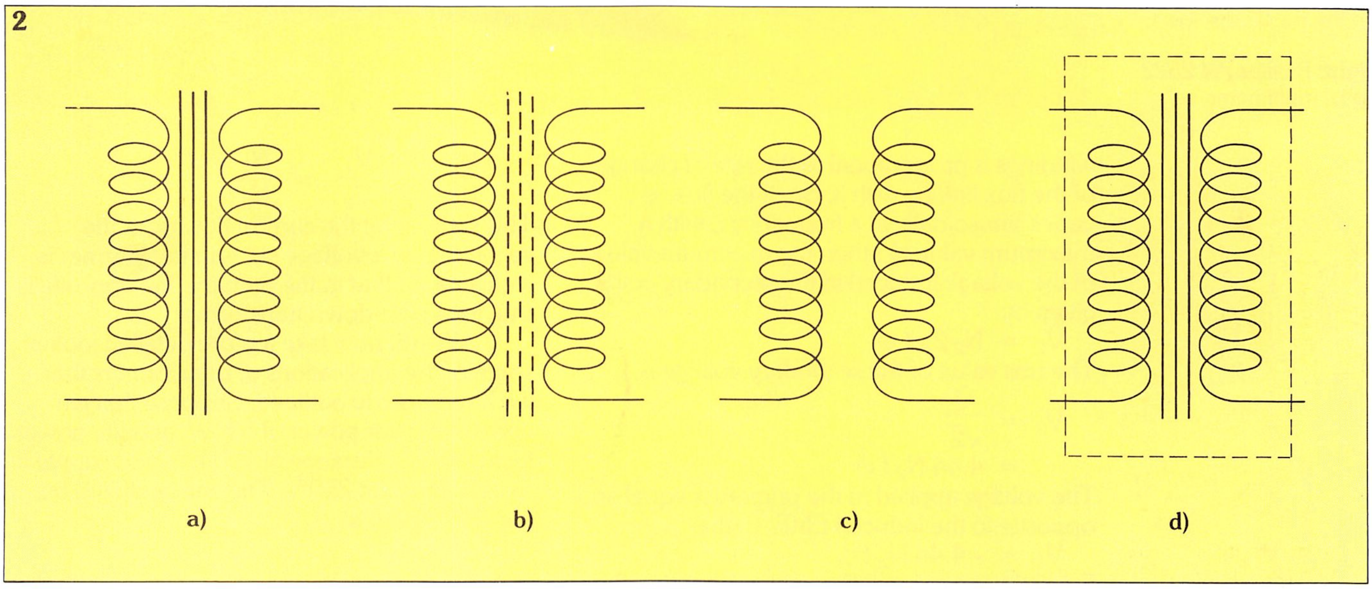
Power dissipation

We stated earlier that the power lost in a transformer was negligible. However, in large transformers, even this small fraction of transmitted power causes considerable heat dissipation within the transformer which needs to be removed to keep the transformer cool. Small transformers operating at less than 100 kVA are adequately cooled by the air. At higher powers, a forced draught may be circulated with a fan, but this is rare.

A more common method of cooling involves immersing the transformer in insulating oil which conducts the heat away. For transformers up to about 30 MVA, the oil can circulate by natural convection and the heat is dissipated by conduction from the surface outside casing – possibly assisted by radiating fins. At even higher power ratings, the oil may be circulated by a pump and cooled, possibly by a heat exchanger.

These factors are applicable to three-phase as well as single-phase transformers, and as electricity supply systems operate on three phases, it is not surprising to find that all large transformers are three-phase devices.

No doubt you have noticed that transformers are rated in VA, rather than the true power transmitted. The insulation in a transformer has to be designed to withstand the maximum voltage on the winding, and as the size of the conductors have to be chosen to carry the maximum current irrespective of the load's power factor, apparent power (the product of voltage and current) is the logical choice to define a transformer's rating.



2. Circuit symbols for transformers.

Circuit symbols

The circuit symbol used to represent a transformer is identical to that used for mutual inductance, and is shown in *figure 2*. The continuous lines between the two coils in *figure 2a* indicate that the transformer is wound onto an iron core; while the broken lines in *figure 2b* indicate that the core material is ferrite. As you may guess, the lack of any lines in *figure 2c* indicates an air cored transformer (which, of course, would be wound onto an insulative former). A dotted box drawn around any transformer symbol (*figure 2d*) indicates that the device is shielded, to prevent it causing electromagnetic interference. □